2.5
2.2
2.0
100 mm
1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm
A5

DOCUMENT RESUME

ED 317 579 TM 014 628

AUTHOR	Graesser, Arthur C.
TITLE	Investigations of Human Question Answering.
INSTITUTION	Memphis State Univ., TN. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Arlington, Va.
REPORT NO	ONR-TR-90-1
PUB DATE	Feb 90
CONTRACT	N00014-88-K-0110
NOTE	80p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Cognitive Processes; *College Students; Higher Education; Interviews; *Models; Prediction; Questioning Techniques; Research Methodology; Surveys; Telephone Surveys; Theories
IDENTIFIERS	*Question Answering; *QUEST Program

ABSTRACT

The development and testing of QUEST, a model of human question answering, are reported. QUEST accounts for answers adults produce for different categories of open-class questions, identifying the information sources associated with the content words in questions. Each information source is organized in a conceptual graph structure. The model suggests three convergence mechanisms that narrow the node space from dozens/hundreds to a handful of nodes that serve as good answers to a question. These mechanisms include: an arc search procedure to restrict the search; answer quality decrease as a function of structural distance; and a constraint satisfaction component to prune out potential answers that are conceptually incompatible with the queried node. QUEST was tested in four informational contexts: (1) expository texts on scientific mechanisms (192 and 162 college undergraduates); (2) narrative texts (110 undergraduates); (3) generic knowledge structures (500 trials with college students); and (4) situations with complex pragmatic constraints (telephone surveys of 336 citizens, interviews with 144 clerks, and filmed interviews using 436 questions). The model successfully predicted the likelihood of generating answers and the goodness-of-answer judgments. Seven tables and 10 figures present study data. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED317579

INVESTIGATIONS OF HUMAN QUESTION ANSWERING

Arthur C. Graesser
Department of Psychology, Department of Mathematical Sciences, and the Institute for Intelligent Systems

Mailing address: Arthur C. Graesser
Department of Psychology
Memphis State University
Memphis, TN 38152
(901) 678-2742

Reproduction in whole or part is permitted for any purpose of the United States Government.

This research was sponsored by the Cognitive Science Program of the Office of Naval Research, under Contract Number N00014-88-K-0110, Contract Authority Identification Number, NR 4422528.

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

TM014628

BEST COPY AVAILABLE

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): ONR TECH REPORT 90-1	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: MEMPHIS STATE UNIVERSITY; 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION: OFFICE OF NAVAL RESEARCH (CODE1142CS)
6c ADDRESS (City, State, and ZIP Code): OFFICE OF SPONSORED PROGRAMS, MEMPHIS, TN 38152	7b ADDRESS (City, State, and ZIP Code): OFFICE OF NAVAL RESEARCH, 800 NORTH QUINCY STREET, ARLINGTON, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-88-K-0110
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR04206	01	4422548

11 TITLE (Include Security Classification): INVESTIGATIONS OF HUMAN QUESTION ANSWERING

12 PERSONAL AUTHOR(S): ARTHUR C. GRAESSER

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
FINAL REPORT	FROM 12/87 TO 12/89	1990, FEBRUARY	70

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	QUESTION ANSWERING, TEXT, NATURAL LANGUAGE COMPREHENSION, COMPUTATIONAL MODELS
05	08		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This project developed and tested a model of human question answering (called QUEST). QUEST accounts for the answers that adults produce when they answer different categories of open-class questions, such as why, how, when, and what-if. QUEST identifies the information sources for questions and assumes that knowledge is organized in the form of conceptual graph structures containing statement nodes and relational arcs. Example types of structures include goal hierarchies, causal networks, taxonomic hierarchies, and spatial partonomies. Question answering procedures operate systematically on these knowledge structures. An important property of QUEST consists of three convergence mechanisms which narrow down the node space from dozens/hundreds of nodes to a handful of nodes that serve as good answers to a question. First, an arc search procedure restricts its search to particular paths of relational arcs, depending on the question category; nodes on legal paths are better answers than nodes on illegal paths. Second, answer quality decreases as a function of structural

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
SUSAN CHIPMAN	202-696-4318	ONR 1142 CS

DD Form 1473, JUN 86 — Previous editions are obsolete — S/N 0102-LF-014-6603

distance (i.e., number of arcs) between the queried node and the answer node. Third, a constraint satisfaction component prunes out potential answers that are conceptually incompatible with the queried node. QUEST also contains a pragmatic component which considers the goals and common ground of the speech participants.

QUEST was tested in the context of expository text on scientific mechanisms, narrative text, and generic concepts. The model successfully predicts the likelihood of generating particular answers to questions and "goodness-of-answer" judgements (and latencies) for particular question-answer pairs. QUEST can also account for answers produced in pragmatically complex contexts, such as telephone surveys, televised interviews, and business transactions.

ABSTRACT

This project developed and tested a model of human question answering (called QUEST). QUEST accounts for the answers that adults produce when they answer different categories of open-class questions, such as why, how, when, and what-if. QUEST identifies the information sources for questions; the primary information sources are associated with the content words in questions (i.e., nouns, main verbs, adjectives). Each information source is organized in the form of a conceptual graph structure that contains nodes and relational arcs. Example types of structures are goal hierarchies, causal networks, taxonomic hierarchies, and spatial partonomies. Question answering procedures operate systematically on these conceptual graph structures.

An important property of QUEST consists of three convergence mechanisms that narrow down the node space from dozens/hundreds of nodes to a handful of nodes which serve as good answers to a question. First, an arc search procedure restricts its search to particular paths of relational arcs, depending on the question category; nodes on legal paths are better answers than nodes on illegal paths. Second, answer quality decreases as a function of structural distance, i.e., the number of arcs between the queried node and answer node. Third, a constraint satisfaction component prunes out potential answers that are conceptually incompatible with the queried node. QUEST also contains a pragmatic component which considers the goals and common ground of speech participants.

QUEST was tested in the context of expository texts on scientific mechanisms, narrative texts, and generic concepts. The model successfully predicted (a) the likelihood of generating particular answers to questions and (b) "goodness-of-answer" judgments for particular question-answer pairs. QUEST can also account for answers produced in conversational contexts that have more complex pragmatic constraints, such as telephone surveys, televised interviews, and business transactions.

INVESTIGATIONS OF HUMAN QUESTION ANSWERING

The purpose of this project was to investigate how adults answer open-class questions, such as why, how, when and what-if questions. We examined the knowledge structures that furnish answers to questions and the cognitive procedures that converge on appropriate and relevant answers to particular questions. This final report on the ONR project has three major parts. First, we provide a brief introduction on question answering research, setting the stage for the research to be reported. Second, we describe QUEST, a model of human question answering that we have developed. Third, we report a series of experiments that test the QUEST model in the context of expository text, narrative text, generic concepts, and naturalistic conversation.

Introduction

Question answering is a very important activity during knowledge acquisition, communication, and social interaction. In spite of this, question asking and question answering have rarely been direct objects of inquiry in the cognitive sciences. Linguists have analyzed the syntactic form of questions, but rarely have attempted to explain the content of answers and the world knowledge that supplies the content. Researchers in education have analyzed the extent to which text acquisition is influenced by adjunct questions, i.e., questions that are placed at the beginning, in the middle, versus at the end of a text, but they have not considered the process of question answering.

Cognitive psychologists have investigated the process of answering closed-class questions at considerable depth (Clark & Clark, 1977; Glucksberg & McCloskey, 1981; Reder, 1982, 1987; Singer, 1986, in press). Appropriate responses to closed-class questions are restricted to a limited number of alternatives and normally are short answers. For example, answers to verification questions are yes, no, maybe, and I don't know. Closed-class questions are contrasted with open-class questions which invite replies with elaborate verbal descriptions, e.g., why, how, when, and what-if questions. Progress in understanding open-class questions has been comparatively slow in psychology (Collins, Warnock, Aiello, & Miller, 1975; Graesser & Black, 1985; Graesser & Golding, 1988; Norman, 1973; Norman & Rumelhart, 1975; Piaget, 1952; Shanon, 1983; Trabasso, van den Broek, & Lui, 1988).

The fields of artificial intelligence and computational linguistics have furnished detailed models of question answering that account for the content of answers and the world knowledge that supplies this content (Allen, 1983; Bruce, 1982; Dahlgren, 1988; Dyer, 1983; Kaplan, 1983; Lehnert, 1978; Lehnert, Dyer, Johnson, Young, & Harley, 1983; McKeown, 1985; Souther, Acker, Lester, & Porter, 1989; Woods, 1977). In most of these models, text and world knowledge are organized in the form of a structured database. Question answering (Q/A) procedures access these information sources and search through the structures systematically. The formalisms and insights from these fields obviously must be tested in psychological experiments before we can incorporate them into psychological models of human question answering. One objective of this ONR contract was to test some of these formalisms and insights.

The development of the QUEST model was influenced by existing Q/A models in cognitive psychology, artificial intelligence, and computational linguistics. We also benefited from available empirical data that was collected in the context of short stories (Goldman & Varnhagen, 1986; Graesser, 1981; Graesser, Robertson, & Anderson, 1981; Graesser & Clark, 1985; Graesser & Murachver, 1985; Trabasso, Stein, & Johnson, 1981; Trabasso et al., 1988), lengthy fairy tales (Graesser, Robertson, Lovelace, & Swinehart, 1980), scripts (Bower, Black, & Turner, 1979; Graesser, 1978), and expository text (Graesser, 1981).

QUEST: A Model of Human Question Answering

It is convenient to segregate QUEST into four major components (see Figure 1). First, QUEST translates the question into a logical form and assigns it to one of several question categories. Second, QUEST identifies the information sources that are relevant to the question. Information sources are represented as conceptual graph structures that contain goal/plan hierarchies, causal networks, taxonomic hierarchies, and descriptive structures. Third, convergence mechanisms compute the subset of nodes in the information sources that serve as relevant answers to a particular question. These convergence mechanisms narrow the node space from hundreds of nodes in the information sources to less than 10 answers to a particular question. Fourth, QUEST considers pragmatic features of the communicative interaction, such as the goals and common ground of the speech participants. Although the process of question answering is segregated into these four components, we acknowledge that an adequate Q/A model would integrate these components in a highly interactive fashion (Dyer, 1983; Lehnert et al., 1983; Robertson, Black, & Lehnert, 1985).

QUEST was not developed to account for the linguistic features of question answering. QUEST does not explain the process of parsing the question syntactically and the process of articulating replies linguistically. Instead, QUEST was developed to account for the conceptual content of the answers.

Question categorization

QUEST assumes that there is a finite set of question categories, that each question category has a unique question answering procedure, and that a particular question is assigned to one of the question categories (see also Lehnert, 1978). For example, How are atoms split? is a "how-event" question which generates causal antecedents to the event "atoms are split." A "why-action" question invites reasons and motives for intentional actions, e.g., Why does a person buy a computer?. A "how-action" question elicits the plan, procedure, and style of executing an intentional action. A "temporal" question elicits the value of the time argument within an event description. QUEST essentially has a catalogue of question categories. Any given question is typically assigned to only one of the question categories.

In order to complete question categorization successfully, it is necessary to determine the node that constitutes the question's "focus" (i.e., the queried node). Several alternative nodes may serve as the question focus in any given question. In the question How is water heated? the question focus is the event "water is heated." This event is a "statement node" which contains a predicate (X heat Y) and an argument (water). Statement nodes are similar to the proposition units which have been frequently adopted as units of representation in the cognitive sciences (Anderson, 1983; Kintsch, 1974; Norman & Rumelhart, 1975). Sometimes the question focus is an argument of a statement node rather than the statement as a whole. In the question What heated the water? the question focus would be X in the statement node "X heated the water."

Some questions are very long-winded and involve many statement nodes: "How does a nuclear power plant in southern California produce electricity when there is a blackout?" In such questions, the focusing mechanism determines which of the alternative nodes is the question focus. The focusing mechanism is complex, with semantic, conceptual, and pragmatic constraints exerting their influences. QUEST does not currently explain the operation of the focusing mechanism; it merely acknowledges this component and assumes that focusing is successfully completed.

Information Sources and Knowledge Representation

An information source is a structured database that furnishes answers to a question. Whenever a question is asked, QUEST computes an expression with three slots:

QUESTION (<Q-category>,<Q-focus>,<information sources>)

The expression for How is water heated? would be:

QUESTION (how-event, water is heated, <information sources>)

The third slot supplies the world knowledge structures that are tapped for answers to the question. At least one information source must be accessed before the question can be completely interpreted and answered. Without an information source, it is difficult, if not impossible, to understand the question and to identify the focus.

When most questions are answered, several information sources are relevant to the question. There are "episodic knowledge structures" (EKSs) that correspond to particular episodes that a person experienced in the past. For example, the answerer might have viewed a film on nuclear power one day, read an article on another day, and had a conversation about nuclear power with a friend on yet another day. These three experiences would create three EKSs in long-term memory. In addition to a large inventory of EKSs shored in memory, there are "generic knowledge structures" (GKSs). A GKS is a more abstract representation that summarizes the typical properties of the content it represents. For example, there are three GKSs triggered by the example question: NUCLEAR-POWER, WATER, and HEAT. The content of the GKS for NUCLEAR POWER would probably include the statement nodes in Figure 2. The content of this GKS is undoubtedly derived from the family of EKSs that are associated with the GKS. However, QUEST does not offer any informative or controversial claims about such relationships. QUEST merely assumes that the cognitive system is a vast storehouse of EKSs and GKSs and that these structures furnish the information sources for questions. Generally speaking, it is easier to access and to search through GKSs than EKSs; GKSs are very familiar knowledge packages that sometimes are products of thousands of experiences.

Many of the information sources for a question are accessed by the content words in the question, such as nouns, main verbs, and adjectives. Information sources that are accessed by content words are called "word-activated" information sources. In contrast, "pattern-activated" information sources are activated by the context of the question and by combinations of content words.

The information sources for a particular question consist of a family of GKSs and EKSs (see Figure 3). Each information source is a structured database with dozens/hundreds of nodes. It follows that there is a wealth of information available in working memory when a question is answered. If there were four information sources and each source had 50 nodes, then 250 nodes would be available. Clearly, most of these nodes would not be produced as answers to the question. Only a small subset of nodes (less than 10) would be produced as answers when adults are asked questions. Convergence mechanisms specify how QUEST begins with 250 possible answers in the node space and converges on approximately 10 good answers.

Graesser and Clark (1985) identified those information sources that are particularly prolific when adults answer questions about episodes in short stories. They reported that word-activated GKSs furnished approximately 72% of the answers whereas pattern-activated GKSs accounted for a modest increment of 8% additional answers. GKSs associated with main verbs in the question were more important information sources than were GKSs associated with nouns. Of course, it is important to acknowledge that these findings may only hold up for simple stories.

There is some debate over the relative contributions of the textbase (an EKS) and generic knowledge structures when questions are answered in the context of text. According to Reder's model of question answering (Reder, 1987), individuals can strategically tap either the textbase or generic knowedge when they decide whether a particular sentence is true or false, and when they decide whether a test sentence had or had not been presented earlier. As Reder articulates it, whenever verification judgments and recognition judgments are made, the person can either access a specific memory (corresponding to the textbase, an EKS) or the person can rely on plausibility judgments (derived from GKSs). As the delay between text comprehension and question answering increases, we rely more on generic knowledge because the textbase is less accessible from memory (Graesser & Nakamura, 1982; Kintsch, 1988; Reder, 1987). When we are unfamiliar with the topic discussed in the text, we rely primarily on the textbase because few if any GKSs are available.

Conceptual graph structures. Most information sources contain a set of units called "statement nodes." According to the representational system adopted by QUEST, the statement nodes in any given information source are organized in the form of a conceptual graph structure. That is, the set of nodes are assigned to node categories and are structured by a network of directed relational arcs. The node categories include state, event, goal, action, and style specification. A state is an ongoing characteristic which remains unchanged throughout the course of the time frame under consideration (e.g., nodes 1,2, and 3 in Figure 2). An event is a state change within the time frame (e.g., nodes 4-10 in Figure 2). A goal refers to an event, state, or style specification that an agent desires (e.g., a person wants to buy a computer, a person wants to be rich). An action is an achieved goal, such that the agent did something that caused the successful outcome). A style specification conveys the speed, intensity, force, or qualitative manner in which an event unfolds (e.g., an event occurs quickly, in circles, quietly). In principle, it is possible to include additional node categories in QUEST, but the above five categories were sufficient for the issues addressed in this project.

There are several categories of arcs in the representational system adopted by QUEST. For the most part, we adopted the arc categories reported in Graesser and Clark (1985). The arc categories in any given information source depend on the type of knowledge depicted in that information source. For example, Consequence (C) arcs are quite prevalent in causal networks (see Figure 2). Goal hierarchies contain the following arc categories: Consequence (C), Implies (Im), Reason (R), Initiate (I), Outcome (O), and Manner (M) arcs. Each arc category is directed, such that the end node is connected to the head of the arc and the source node is connected to the tail:

(source node) ---ARC---> (end node)

Table 1 presents the rules of composition for these six arc categories. A complete description of this representational system is beyond the scope of this report (see Graesser & Clark, 1985). The examples in this report should convey the important characteristics of the representational system.

There are semantic and conceptual constraints that must be satisfied before two nodes can be connected by an arc of a particular category. For example, there are three constraints associated with the Consequence arcs which are prevalent in Figure 2. Consequence arcs can relate event/state/style nodes but not goal nodes. The source node must have occurred or existed in time prior to the end node. That is, the cause must precede the effect. The source node must play some causal role in producing the end node (Trabasso et al., 1988). That is, if the source node is negated or removed, then the end node would never occur. Two nodes cannot be related by a Consequence arc if any one of these three constraints are violated.

Conceptual graph structures have foundations in a number of representational systems in the cognitive sciences. These systems include propositional theories (Anderson, 1983; Clark & Clark, 1977; Kintsch, 1974; Norman & Rumelhart, 1975), story grammars (Mandler, 1984; Stein & Glenn, 1979; Trabasso, Stein, & Johnson, 1981), causal chain theories (Black & Bower, 1980; Trabasso & van den Broek, 1985; Trabasso et al., 1988), conceptual dependency theory (Alterman, 1988; Schank & Abelson, 1977; Schank & Reisbeck, 1981), rhetorical organization (Mann & Thompson, 1986; Meyer, 1985), and conceptual graphs (Sowa, 1983).

Types of knowledge structures. We have devoted most of our research efforts on four types of knowledge structures. First, causal networks (as illustrated in Figure 2) contain event chains, along with states that enable the events. Second, goal hierarchies convey the plans and intentional actions that are executed by animate agents, along with states/events in the world that trigger these goal hierarchies (see Figure 4). Third, taxonomic hierarchies specify how classes of entities are nested hierarchically within other classes (see Figure 5, top). Fourth, spatial partonomies specify how regions are embedded within other regions, along with the relative positions of regions (see Figure 5, bottom). We acknowledge that a particular information source is an amalgamation of all four types of structures. The purpose of segregating these types of structures is to identify the systematic characteristics of both the structures and the Q/A procedures that operate on the structures.

The research in this project concentrated primarily on the causal networks and goal hierarchies so this report will hereafter focus on these two types of structures. Graesser and Franklin (in press) has a more detailed description of QUEST in the context of all four types of knowledge structure.

Convergence Mechanisms

When a particular question is asked, QUEST activates several information sources in working memory and each information source has dozens/hundreds of nodes. Convergence mechanisms narrow down the node space from hundreds of nodes to approximately 10 good answers. Convergence is accomplished by three components: (1) an intersecting node identifier, (2) an arc search procedure, and (3) constraint satisfaction.

Intersecting nodes and structural distance. An intersecting node identifier isolates those statement nodes from different knowledge structures that intersect (i.e., match, overlap). For example, the statement node "electricity is produced" may be stored in several information sources within working memory. There is evidence that these intersecting nodes have a higher likelihood of being produced as answers than do nonintersecting nodes (Golding, Graesser, & Millis, in press; Graesser & Clark, 1985). In addition, nonintersecting nodes have a lower likelihood of being produced as answers to the extent that they are more arcs away from an intersecting node. The likelihood of a node being produced as an answer decreases exponentially as a function of its structural distance from the nearest intersecting node (Graesser & Clark, 1985; Graesser & Hemphill, 1990; Graesser, Hemphill, & Brainerd, 1989).

The bias toward intersecting nodes provides one convergence mechanism but does not go the distance in reducing the node space from hundreds of nodes to 10 nodes. The arc search procedure and constraint satisfaction reduce the node space even further. It is important to note that the reported impact of structural distance on answer production does partial out contributions from the arc search procedure and constraint satisfaction.

The predicted effects of structural distance on answer production and answer quality would be generated by some models of question answering other than QUEST (Shanon, 1983; Winston, 1984). Winston has described a Q/A model that answers why and how questions in the context of goal/plan hierarchies and problem spaces. His Q/A algorithm specifies that good answers are only

one arc away from the queried node. Theories of marker passing and spreading activation (Anderson, 1983) would also predict the distance gradient.

Arc search procedure. Each question category has its own arc search procedure that operates on the information sources relevant to a question. When a given information source is accessed, the arc search procedure first identifies an "entry node" in the information source. The entry node usually matches the question focus. For example, if the question is How is water heated? and the information source is Figure 2, then event 6 would be the entry node in the structure. In some cases, the entry node does not match the question focus; instead it matches an intersecting node between two information sources. For example, the node "energy is release" might be stored in the GKS for NUCLEAR-POWER and the GKS for HEAT. These intersecting nodes would serve as entry nodes even though they do not match the question focus ("water is heated").

Once an entry node is located in an information source, the arc search procedure executes a breadth-first search from the entry node by pursuing legal arcs that radiate from the entry node. For each question category, there is a particular set of arc categories and arc directions that are legal. Figure 6 shows the legal paths for queried events. Some question categories pursue causal antecedents on paths of backward Consequence arcs (namely the why, how, when, and enable questions) whereas other question categories pursue causal consequences via forward Consequence arcs (namely consequence and what-if questions). Consider the question How is water heated? in the context of Figure 2. Legal answers would be nodes 1, 2, 4, and 5 whereas illegal answers would be nodes 3, 7, 8, 9, and 10. The legal answers would be entirely different for the question What are the consequences of water being heated?: nodes 7-10 but not nodes 1-5. The fact that illegal paths are pruned from consideration substantially narrows down the node space.

Complex knowledge structures have a more diverse distribution of arc categories than merely Consequence arcs. A complete specification of a causal antecedent path consists of any combination of the following arcs: Implies (forward or backward), backward Outcome, backward Initiate, and backward Consequence. A causal consequence path consists of any combination of the following arcs: Implies, forward Initiate, forward Outcome, and forward Consequence. A complete account of the legal paths for different question categories is provided in previous publications (Graesser & Clark, 1985; Graesser & Franklin, in press; Graesser & Murachver, 1985).

Goal hierarchies (see Figure 4) contain a set of goal nodes that are interrelated by Reason arcs. Goal hierarchies frequently have Manner arcs and other characteristics but we will ignore these for the moment. Superordinate goals are at the top of the goal hierarchy whereas low-level subordinate goals and actions are at the bottom of the hierarchy. Figure 6 shows the arc search procedures when a goal (or action) node is probed with a question. Answers to why and consequence (abbreviated as CONS) questions pursue superordinate goals via forward Reason arcs. Answers to how, when, and enable questions tend to pursue subordinate goals that radiate from the entry node on paths of backward Reason arcs. For example, consider the question Why did Bill call Jill on the telephone? Legal goal answers would be nodes 1 and 2 (in order to feel better, in order to talk to Jill) whereas illegal goal answers would be nodes 3, 5, and 6 (in order to go to a bar, in order to walk to a couch, in order to dial Jill's number). The legal goal answers would be entirely different for the question How did Bill call Jill on the telephone? The legal answers would be nodes 5 and 6 (Bill walked to a couch, Bill dialed Jill's number) but not nodes 1, 2, and 3.

Goal-oriented knowledge is normally more complex than a structure with goal nodes interconnected by Reason arcs (Miller, Gallanter, & Pribram, 1960; Newell & Simon, 1972; Schmidt, Sridharan, & Goodson, 1978; Wilensky, 1983).

(1) Sets of goal nodes are frequently packaged in the form of plans or scripts (Schank & Abelson, 1977).
(2) Some goal nodes are connected by Manner arcs rather than Reason arcs.
(3) Goal nodes are normally triggered by states/events in the world by virtue of the Initiate arcs. For example, nodes 7-10 initate the goal hierarchy in Figure 4 (nodes 1-6).
(4) Sibling nodes in a goal hierarchy are related by additional arcs (and, or, before). Sibling nodes are immediately dominated by the same parent goal. For example, nodes 3 and 4 in Figure 4 would be related by or because the achievement of either one of these goals would end up enabling the parent node. Nodes 5 and 6 would be related by a before arc because goal 5 would need to be achieved prior to goal 6.
(5) Goal nodes may or may not be achieved when a goal hierarchy is instantiated. When a goal node is achieved, an event/state is constructed and linked to the goal node by an Outcome arc. For example, if Bill manages to dial Jill's number, then there would be an event node (Bill dialed Jill's number) in addition to the goal node (Bill wanted to dial Jill's number). An intentional action is an amalgamation of a goal and its outcome node. When a goal is not achieved, there either is no goal node or a negative outcome node (e.g., Bill did not dial Jill's number).

Given that goal hierarchies are rather complex structures, the arc search procedures for queried actions are more complex than the procedures depicted in Figure 6. For example, answers to why questions include (a) superordinate goals in the goal hierarchy, (b) sibling nodes that precede the entry node, (c) states/events that initiate the goals, and (d) causal antecedents to the goal initiators. Once again, a more complete specification of the arc search procedures are provided in previously published studies (Graesser & Clark, 1985; Graesser & Franklin, in press; Graesser & Murachver, 1985).

We have written a computer program that generates legal answers to many different types of questions that may be asked in the context of causal networks, goal hierarchies, taxonomic structures, and spatial partonomies. The user specifies one or more information sources and then enters the question. The computer lists all answers that would pass the arc search procedure associated with the question category. The program is written in Common LISP. It has been implemented on a microcomputer (IBM clone), a LISP machine (Texas Instrument EXPLORER II), and a parallel computer (INTEL hypercube with 16 parallel systems). We refer to these implementations as microQUEST, QUEST, and hyperQUEST, respectively. QUEST and hyperQUEST are needed whenever multiple information sources are relevant to a question.

The arc search procedures in QUEST are compatible with some theories of question answering in artificial intelligence which have emphasized the importance of knowledge organization and of restricting search by pursuing particular conceptual relations (Dahlgren, 1988; Lehnert, 1978; Lehnert et al., 1983; Schank & Abelson, 1977; Souther et al., 1989; Winston, 1984).

Constraint satisfaction. The semantic and conceptual content of the answer should not be incompatible with the content of the queried node. Constraint satisfaction discards those candidate answers in the node space which are incompatible with the focus slot of the question. Stated differently, the question focus has semantic and conceptual constraints that are propagated among nodes in the node space, ultimately pruning out the incompatible nodes.

There are several ways in which a candidate answer could be incompatible with the question focus. Dimensions of incompatibility have been identified and confirmed by Graesser and Clark (1985). The example answers below illustrate two of these dimensions in the context of the question focus water is heated and a nuclear power plant.

(1) The water is frozen. This node directly contradicts the question focus.

(2) Water fell from the clouds. This node involves a "time frame" incompatibility because rainfall for a given water molecule is entirely outside of the time frame of the same water molecule being heated in a nuclear power plant.

Contradiction and time frame incompatibility hardly exhaust the possible dimensions that would be computed during constraint satisfaction. "Planning incompatibilities" occur whenever a plan conveyed in a candidate answer (e.g., person saves money) is incompatible with the plan conveyed in the queried node (e.g., person buys computer). "Causal strength" is the extent to which the candidate answer is causally related to the queried node; a candidate node would be pruned out if it is not causally related to the queried node. "Argument overlap" computes whether the two nodes (candidate answer and queried node) share one or more common arguments. A "plausibility" dimension specifies whether the candidate answer is true or false with respect to general world knowledge; implausible nodes in the knowledge base would be pruned out.

A simple computation of constraint satisfaction would evaluate each candidate node on all dimensions: contradiction, time frame incompatibility, planning incompatibility, causal strength, argument overlap, and plausibility. According to a strict criterion, a candidate node passes constraint satisfaction if all of the dimensions are satisfied. According to a weak criterion, a node passes if most dimensions are satisfied or if most dimensions are satisfied to some degree.

Interactions among components of convergence. The three components of convergence (node intersection & structural distance, arc search procedure, and constraint satisfaction) are able to narrow the space of candidate nodes to approximately 10 good answers to a question. Given the adequacy of these convergence mechanisms, one might then consider how the three components interact. For example, one position might be that the three components operate independently and additively. If so, there should be main effects but no interactions in analyses of answer quality and in analyses of answer production scores. Alternatively, the three components might be executed interactively. Of course, this latter position would not be particularly informative unless it predicted systematic output.

One obvious position to consider is that the components are executed in sequential order, with operation of component N being dependent on the output from component N-1. For example, suppose that the process sequence below was correct.

Stage 1. Find all intersecting nodes among the information sources.

Stage 2. For each information source, find an entry node (e.g., corresponding to the question focus) and randomly sample a candidate answer node.

Stage 3. Apply the arc search procedure to determine whether the candidate answer is on a legal path extending from the entry node. If the answer is illegal, then stop and decide that the candidate node is a bad answer.

Stage 4. Apply the constraint satisfaction component to check whether the candidate answer node is compatible with the entry node. If it is not compatible, then stop and decide that the candidate node is a bad answer.

Stage 5. Compute the structural distance between the candidate node and the entry node. If the distance score is high, then decide that the candidate node is a bad answer. If the distance score is low, then decide that the candidate node is a good answer.

The above sequential processing mechanism would predict some interactions. First, there would be a 2-way interaction between arc search and constraint satisfaction; constraint satisfaction would predict answer quality for legal answers (which pass onto stage 4) but not illegal answers (which do not reach stage 4). Second, there would be a 3-way interaction among arc search, constraint satisfaction, and structural distance; structural distance would influence answer quality scores only for legal answers that satisfy constraints (i.e., nodes that reach stage 5). Decision latencies would generally be longer for legal than illegal answers because legal answers would pass through more stages.

At this point, the QUEST model does not specify particular interactions among convergence components. There simply is not enough empirical data to make any definitive claims. However, the studies in this report do periodically provide some data that are relevant to this issue.

Pragmatics

The pragmatic components address the social and communicative functions of answering a question. One component considers the goals of the questioner and answerer. From the perspective of the questioner, a question may be asked in order to acquire information, to solve a problem, to assess how much the answerer knows, to persuade, to control a conversation, and so on. From the perspective of the answerer, the answer may be formulated to inform the questioner, to let the questioner know the answerer knows something, to entertain the questioner, and so on. A complete model of Q/A would consider the goals of the speech participants in a particular discourse context and would determine how the answers are tailored to achieve these goals (Allen, 1983; Appelt, 1985; Bruce, 1982; Francik & Clark, 1985; Kaplan, 1983).

One important goal to assess is whether the questioner genuinely seeks the information suggested by the question. Some questions are not genuine information seeking questions: Indirect requests (e.g., Would you pass the salt?), greetings (How are you doing?), gripes (Why does this always happen to me?), and rhetorical questions. Van der Meij (1987) has identified the assumptions that must be met before an utterance constitutes a genuine information seeking question. These assumptions are listed below.

(1) The questioner does not know the information asked for with the question.
(2) The questioner believes that the presuppositions of the question are true.
(3) The questioner believes that an answer exists.
(4) The questioner wants to know the answer.
(5) The questioner can assess whether a reply constitutes an answer.
(6) The questioner poses the question only if the benefits exceed the costs. For example, the benefits of knowing the answer must exceed the costs of asking the question.
(7) The questioner believes that the answerer knows the answer.
(8) The questioner believes that the answerer will not give an answer in absence of the question.
(9) The questioner believes that the answerer will supply an answer.

A second pragmatic component is the common ground (i.e., shared knowledge, mutual knowledge) between questioner and answerer (Clark & Marshall, 1981; Miyake & Norman, 1979; Shanon, 1983; Sleeman & Brown, 1982). According to these models, the answerer first estimates the common ground between speech participants and then selects an answer that moderately extends the boundaries of the common ground. That is, the answer should be somewhat more informative, elaborate, or detailed than the common ground, but should not be (a) entirely within the sphere of the common ground or (b) substantially more detailed than the common ground.

In principle, the QUEST model is able to keep track of the common ground between questioner and answerer. QUEST would evaluate what information sources the questioner has stored in memory and what nodes the questioner has stored in each information source. The fringe or boundary of knowledge can also be computed in a straightforward manner. In particular, a fringe answer would be few arcs away from a node in the common ground.

Common ground could have some counterintuitive effects on answer quality. If the common ground is high and the answerer wants to be informative (i.e., supplying information that the questioner might not know), then the answerer would avoid nodes that are in multiple information sources. Surprisingly, there would be a negative correlation between answer quality and number of information sources. Regarding structural distance and common ground, there might be a preference for distant nodes because proximate nodes would be easy to infer. Perhaps a curvilinear relationship would occur, with answers at intermediate distances being better than answers at close and at far distances from the entry node in an information source.

This section has described QUEST and has identified its theoretical foundations. One of the objectives of the ONR contract was to assess the extent to which QUEST's components can explain empirical data in question answering tasks. The studies reported in the next section were completed under the ONR contract in order to address this objective. The results of these studies were quite promising. Therefore, we conclude that QUEST is a plausible psychological model of human question answering.

Tests of the QUEST Model

QUEST was tested in four different informational contexts. These contexts included (1) expository texts on scientific mechanisms, (2) narrative texts, (3) generic knowledge structures (e.g., objects, person concepts, scripts), and (4) situations with complex pragmatic constraints (i.e., telephone surveys, business transactions, and televised interviews). In some experiments, we examined answer production scores, that is, the likelihood that a node in an information source was produced as an answer to a question. In other experiments, we examined answer quality scores for question-answer pairs, i.e., whether the answer is a good versus a bad answer to the question.

Studies of Expository Texts on Scientific Mechanisms

Two studies focused on short texts that describe event chains in physical, biological, and technological systems (Graesser & Hemphill, 1989; Graesser, Hemphill, & Brainerd, 1989). Each text had five events, as illustrated in the text below on nuclear power.

1. Atoms are split into particles.
2. Heat energy is released.
3. Water in the surrounding tank is heated.
4. Steam drives a series of turbines.
5. The turbines produce electricity.

We assumed that college students had very little knowledge about these scientific mechanisms and that they relied primarily on the textbase as an information source. Moreover, we started out with the simple assumption that the textbase consisted of a linear chain of events, connected by Consequence arcs.

[E1] --C--> [E2] --C--> [E3] --C--> [E4] --C--> [E5]

We eventually had to revise this simple assumption because readers sometimes imposed a goal-oriented, teleological interpretation on the event sequences in the case of technological and biological mechanisms. That is, one event occurred for the purpose of achieving subsequent events, e.g., water is heated (event 3) for the purpose of having steam drive a series of turbines (event 4). Indeed, the engineers of a nuclear power plant would design the plant with such goals in mind. Whenever a teleological interpretation was imposed on the text, the textbase consisted of a goal structure running parallel with the causal chain (see Figure 7).

We tested QUEST by examining the answers to five question categories: why, how, when, enable, and consequence (CONS). For example, event 3 would be probed with the following questions:

Why is water heated?
How is water heated?
When is water heated?
What enabled water to be heated?
What are the consequences of water being heated?

Given that each text had 5 events and that there were 5 question categories, 25 unique questions were associated with each text. Although subjects occasionally generated inferences when they answered these questions, we analyzed only those answers that referred to events explicitly stated in the text. Therefore, we were concerned with four possible answers to each question. For example, the queried node in the above questions is event 3; we analyzed those answers that referred to events 1, 2, 4, and 5.

These studies on expository text were designed to test the three components of the convergence mechanism: Structural distance, the arc search procedures, and constraint satisfaction. We found robust support for structural distance and the arc search components but not for constraint satisfaction. Therefore, we will concentrate primarily on the two successful components; analysis of constraint satisfaction will be saved for the end of this subsection.

Figure 6 shows the arc search procedures of the QUEST model. Legal answers to how, enable, and when questions are causal antecedents to the queried event. If event 3 were probed with these types of questions, then legal answers would be events 1 and 2 but not events 4 and 5. Legal answers to CONS questions are causal consequences (events 4 and 5 but not events 1 and 2). Legal answers to why questions depend on whether a goal structure is superimposed on the causal chain. If not, then legal answers to why questions are the same as answers to how, enable, and when questions. If a goal structure is superimposed, however, then legal answers include causal consequences but not causal antecedents.

The structural distance component predicts that proximate answers should be better than distant answers. Presumably, structural distance has an influence on legal answers but not on illegal answers. If event 3 were probed with the questions, then the following predictions would be generated by QUEST.

How, enable, when, and why(nonteleological) questions

$$E2 > E1 > E4 = E5$$

CONS and why(teleological) questions

$$E4 > E5 > E1 = E2$$

Given that all 5 events in a text were queried and that there are 4 possible answers per question, there are a total of 20 cells in a complete "question-answer matrix." We analyzed how well QUEST could account for the dependent measures (i.e., answer production scores and answer quality scores) in the question-answer matrices. A question-answer matrix was prepared for each text, and separate matrices were prepared for each question category.

Texts. The texts were 24 event sequences that were extracted from passages in the American Academic Encyclopedia. All texts had five events, as in the example about nuclear power. Eight passages were in the technological domain (computer, television, paper production, nuclear energy, elevator, vacuum cleaner, water purification, and wine production); eight were in the biological domain (heart, seeing, photosynthesis, knee jerk, mitosis, hair growth, hearing, neurons); and 8 were in the physical science domain (tornado, earthquake, light, rain, sonic boom, riptide, supernova, and stalagmites).

A question-answer matrix was prepared for each text and each question category in all analyses of answer quality scores and answer production scores. Given that there were 24 texts, 5 question categories, and 20 cells per question-answer matrix, 2400 scores were included in the item analyses.

Answer production scores. We collected answer production scores from 192 undergraduate students at Memphis State University. The subjects first read one of the 24 texts and later answered 25 questions about the text (5 events x 5 question categories). The 25 questions were randomly presented in a booklet. Two blank lines appeared after each question for subjects to write down their answers. Eight subjects were randomly assigned to each of the 24 texts.

The critical dependent measure was the answer production score. This was computed as the proportion of subjects (out of 8) who generated a particular answer to a particular queried event. The score associated with a particular question-answer item was the basic unit in all quantitative analyses. Therefore, all tests of statistical significance assessed variability among items in its error term, but not variability among subjects. All statisitical tests were performed at the $p < .05$ level. (We will not present the exact F-scores in this report; instead, we will simply announce whether an effect was versus was not significant.)

Figure 8 presents question-answer matrices for the five question categories. Each matrix presents answer production scores as a function of the five serial positions for questions and the five serial positions for answers. The scores in the upper-right half of each matrix correspond to causal consequences whereas the scores in the bottom-left half of the matrix correspond to causal antecedents.

An inspection of the mean answer production scores confirm QUEST's arc search procedures for how, when, enable, and CONS questions. Mean scores of causal antecedents were significantly higher than the scores of causal consequences in the case of how questions (.21 versus .06), when questions (.36 versus .10), and enable questions (.26 versus .08); as predicted, CONS questions showed the opposite pattern (.08 versus .36). However, mean scores of the why questions were approximately the same for causal antecedents and causal consequences (.16 versus .17).

As discussed earlier, answers to why questions would tap causal consequences if teleological structures (i.e., goal structures) were superimposed onto the causal chains (see Figure 7). We would expect these goal structures to be imposed on technological domains and perhaps biological domains, but not on physical science domains. Therefore, we performed separate analyses on the three types of domains (see Figure 8). When why questions were asked in the context of physical science, mean scores were significantly higher for causal antecedents than causal consequences (.30 versus .10). Thus, goal structures were not constructed when these physical science texts were comprehended. In contrast, mean scores were significantly lower for causal antecedents than for causal consequences when why questions were asked about events in technology (.08 versus .20) and in biology (.10 versus .20). A teleological goal structure was the primary structure in the case of technological and biological domains.

QUEST's prediction about the impact of structural distance on answer production scores was also confirmed. This should be apparent when inspecting the scores in Figure 8. The scores decreased as a function of the distance (number of arcs) between the queried node and the answer node. When averaging over all five question categories, the mean answer production scores for legal answers significantly decreased as a function of distance, .42, .28, .21, and .20 at distances of 1, 2, 3, and 4, respectively. This decrease fit an exponential function better than a linear function. In contrast, structural distance did not have a consistent significant effect on illegal answers. When averaging over the five question categories, the scores were .12, .06, .04, and .07 for distances of 1, 2, 3, and 4, respectively.

A very simple mathematical model with three parameters closely fit the answer production scores. Parameter a is the likelihood of pursuing a causal antecedent path whereas parameter c is the likelihood of pursuing a causal consequence path. Parameter t is the likelihood of traversing a single arc on a path. There is fixed parameter n that consists of the number of arcs between the queried node and the answer node. The prediced answer production scores are computed as $a*t^n$ for causal antecedents and $c*t^n$ for causal consequences. This simple model accounted for 89% of the variance of the answer production scores. The best fit value of t was .67, the distance dampening parameter. Regarding the causal antecedent parameter (a), the best-fit value was substantially higher in those conditions in which antecedents were legal (.63) than in those conditions in which consequences were legal (.18). Regarding the causal consequence parameter (c), again the best-fit values were higher when consequences were legal (.51) than when consequences were illegal (.19).

Goodness-of-answer judgments. Goodness-of-answer (GOA) judgements were collected on the 24 texts and were analyzed in the same way as the answer producted scores reported above. After the subjects read a text, they were presented a series of question-answer pairs. On each of these trials the subject decided whether the answer was a good versus a bad answer to the question. The patterns of GOA judgments were expected to be similar to those of the answer production scores even though the tasks were somewhat different. Whereas the answer production task requires the subject to retrieve answers from memory, the GOA task places less demands on memory and more demands on the judgment of answer quality. QUEST makes idential predictions across tasks regarding the arc search and structural distance components.

The GOA task permitted us to impose some control over the particular arc categories pursued when the arc search procedures are executed. This control is achieved by specifying the connective that precedes the answer. In the case of why-questions, causal antecedent paths should be pursued when the answer is preceded by because; that is, backward Consequence arcs are sampled. In contrast, when the answer is preceded by in order to/for, the goal structure and forward Reason arcs should be sampled, corresponding to causal consequences. For

example, consider the following four answers to the question Why is water heated? in the context of the nuclear power text.

Because atoms are split. (causal antecedent)
* In order for atoms to be split. (subordinate goal)
* Because turbines produce electricity. (causal consequence)
In order for turbines to produce electricity. (superordinate goal).

The answers with asterisks are illegal whereas the answers without asterisks are legal, according to QUEST.

Legal answers to when-questions are also governed by connectives. If the answer is preceded by after, then legal answers are causal antecedents. If the answer is preceded by before, then legal answers are causal consequences. If no connective precedes the answer, then legal answers are causal antecedents, according to QUEST. That is, the default connective is after (see Graesser & Franklin, in press; Graesser & Murachver, 1985).

In light of the expected impact of connectives on GOA judgements, this study had nine conditions altogether. In five of the conditions there were no connectives preceding the answer, corresponding to the five types of questions. There were four conditions with connectives preceding answers. There were two connective conditions for why-questions (because versus in order to/for) and two for when-questions (after versus before).

The subjects were 162 undergraduates at Memphis State University. These subjects were randomly assigned to the nine question conditions, with 18 subjects per condition. Half of the subjects read and were tested on 12 of the 24 texts; the other half of the subjects were assigned the other 12 texts. The presentation order of the texts was randomized for each subject. Associated with each 5-event text was 20 different question-answer items, as delineated in the 20-cell question-answer matrix. After the subject read a text, the subject provided GOA judgments on all 20 items, which were presented in random order.

A microcomputer controlled the presentation of the passages, the question-answer items, and the collection of responses. The subject began each question-answer trial by pressing one of the keys upon receiving a READY signal. After a .5 second delay, the question appeared on the screen. The subject read the question at his own pace and pressed a key when finished. After a .5 second delay, the question disappeared and the answer appeared on the screen. The subject provided the GOA judgment by pressing one of two response buttons (GOOD versus BAD).

A GOA score was computed as the proportion of observations in which subjects judged an answer as GOOD. In all tests of statistical significance we were able to compute a minF statistic; this is a conservative test that considers both variability among subjects and variability among items. As in Figure 8, we prepared a question answer matrix for each text, for each subject, and for each question category.

Mean GOA judgments were in the expected directions when comparing causal antecedents and causal consequences. GOA judgments were significantly higher for antecedents than for consequences in the following six conditions: why [.48 versus .29], why(because) [.52, .27], how [.48, .18], when [.53, .21], when(after) [.67, .13], and enable [.63, .23]. GOA judgments were significantly lower for causal antecedents than for causal consequences in the following three question groups: why(in order) [.26, .57], when(before) [.16, .79], and CONS [.25, .55]. The difference between antecedents and consequences was smallest in the why group (.19); this would be expected because this is the only group in which legal answers depend on the knowledge domain (i.e., physical science, biological, versus technological).

The GOA judgments in the 20-cell matrices also showed effects of structural distance. The legal answers showed an exponential decrease as a function of structural distance, .76, .52, .42, and .39 for distances of 1, 2, 3, and 4, respectively (averaging across question categories). The differences were very subtle for illegal answers, .27, .19, .17, versus .19.

Constraint satisfaction and multiple regression analyses. The patterns of GOA judgments in the above analyses confirmed QUEST's predictions with respect to the arc search procedure and structural distance. We also performed analyses that assessed the impact of constraint satisfaction on GOA judgments. Two dimensions were considered in our analyses of constraint satisfaction: Argument overlap and causal strength. Presumably, answers would have a higher GOA judgment if the answer and the queried node shared at least one noun-argument and if there was a high causal strength between the two nodes.

We adopted van den Broek and Trabasso's analysis of causality when we scaled event pairs on causal strength (van den Broek, 1990; Trabasso & van den Broek, 1985; Trabasso et al., 1988). Causality is decomposed into four criteria: temporality, operativity, necessity, and sufficiency. The temporality criterion states that the event X (the cause) must precede event Y (the effect) in time. The operativity criterion states that event X or the result of event X must be operating when event Y occurs. The necessity criterion is satsified if events X and Y pass the counterfactual test; that is, event X is necessary for event Y if event Y fails to occur when event X is negated. According to the sufficiency criterion, X is sufficient for Y under the following conditions: If event X occurs and normal circumstances in the world continue, then event Y will occur. Each of these four criteria received a value ranging from 0 to 1 whenever two events were evaluated on causality. The overall causal strength between X and Y was computed according to formula 1.

$$\text{Causal strength} = T^{*}O^{*}(N+S)/2 \qquad (1)$$

T, O, N, and S refer to the values of temporality, operativity, necessity, and sufficiency, respectively. Trained judges rated event pairs on these four criteria (i.e., assigning values of 0, .5, or 1) and achieved a high level of agreement (.70 or more decisions being the same for any pair of judges).

We performed multiple regression analyses in order to assess the impact of several variables on answer production scores (or on GOA judgments) for question-answer pairs. First, there was the arc search variable, which had values of 0 (illegal answer) or 1 (legal answer). Second there was structural distance, the number of arcs between the queried node and the answer node. Third, there was causal strength, as measured above, which varied from 0 to 1. Fourth, there was argument overlap, which had values of 0 (no arguments overlap) and 1 (at least one argument overlaps). Fifth, there was topic familiarity. Millis (1989) collected familiarity ratings from college students on the 24 texts; the values of familiarity ranged from 1 (very unfamiliar with the topic) to 6 (very familiar with the topic). Sixth, there was knowledge domain, with dummy coded variables corresponding to the technological, biological, and physical science domains.

In analyses of how, when, enable, and CONS questions, knowledge domain virtually never interacted with any of the other predictor variables, so we will not report separate regression analyses on each knowledge domain. However, for why questions, there were frequent interactions between knowledge domain and the arc search procedure, so separate regression analyses will be reported on technological, biological, and physical science domains.

Table 2 presents results of the multiple regression analyses on answer production scores. In all 7 regression analyses, the overall multiple regression equation significantly predicted answer production scores. The mean percentage of variance explained by the equation was 32% (whe weighting the five question categories equally). Standardized beta-weights are presented for each predictor variable, along with an indication of whether the predictor had a significant

semipartial correlation. All multiple regression analyses showed significant effects of arc search (mean beta = .44) and structural distance (mean beta = -.24). Causal strength was significant in only 1 out of the 7 analyses (mean beta = .04) and argument overlap was significant in 2 of the 7 analyses (mean beta = .05). In 4 of the 7 analyses, the answer production scores decreased significantly as a function of topic familiarity (mean beta = -.08).

In a set of follow-up regression analyses we added interaction terms for three components of the convergence mechanism, namely arc search (A), structural distance (D), and causal strength (C). There were four possible interaction terms: AxC, AxD, CxD, and AxCxD. The inclusion of the interaction terms increased the amount of explained variance from 32% to 57%. The causal strength variable rarely interacted significantly with the other two convergence components; the AxC, DxC, and AxCxD interactions were each significant in only 1 out of 7 multiple regression analyses. The AxD interaction was the only robust and consistent interaction (6 out of 7 analyses). This AxD interaction reflected the pattern of data reported earlier. That is, structural distance had a large impact on legal answers but a very subtle impact on illegal answers.

Multiple regression analyses were also performed on the GOA judgments, using the same statistical procedures that were used on the answer production scores. All multiple regression analyses significantly predicted the GOA judgments; the mean percentage of variance explained by the multiple regression equations was 57% (varying from 30% to 85%). Table 3 presents beta-weights for arc search, structural distance, and causal strength. The arc search variable was significant in all 15 analyses (mean beta = .57) whereas structural distance was significantly negative in 14 out of 15 analyses (mean beta = -.36). Causal strength was significant in only 2 out of 15 analyses (mean beta = .02). In addition, topic familiarity was significantly positive in 2 analyses (mean beta = -.03) and argument overlap was significant in 3 out of 15 analyses (mean beta = -.03).

In a set of follow-up multiple regression analyses, we added interaction terms for arc search, structural distance, and causal strength (AxC, AxD, CxD, and AxCxD). The AxD interaction was significant in most of the analyses (9 out of 15) as was the case in the analyses of answer prodution scores. Once again, the causal strength predictor rarely interacted with the other two convergence components; the AxC, CxD, and AxCxD interactions were significant in only 2 out of 45 cases.

To summarize, these multiple regression analyses provided consistent and robust support for the arc search procedures and for structural distance, but failed to show effects of causal strength and argument overlap (i.e., two dimensions of constraint satisfaction). We have a plausible explanation for the finding that causal strength had no impact on answer production scores and GOA judgments. Individuals may need a sufficiently deep level of understanding about the topic before they can construct causal interpretations of the events. That is, an analysis of temporality, operativity, necessity, and sufficiency requires a rich body of world knowledge. The college students probably had a very superficial understanding of the 24 texts so effects of causal strength failed to emerge.

Decision latencies for GOA judgments. We analyzed the decision latencies of the GOA judgments in the above experiment. The mean decision latencies varied from 2.16 seconds for CONS questions to 3.16 seconds for when(before) questions. Table 3 shows the outcome of the multiple regression analyses on the decision latencies. The regression equation significantly predicted latencies in all 15 equations and accounted for 12% of the item variance.

Table 3 shows beta-weights for the arc search, structural distance, and causal strength predictors. Arc search was significantly positive in 8 out of 10 analyses in which legal answers were antecedents (mean beta = .23); in these cases, decision times were longer for legal answers than for illegal answers. Arc search was not significant in any of the analyses in which consequences

were legal answers, i.e., CONS, when(before), and 3 why(in order). Structural distance was significantly negative in 7 out 15 analyses; the sign was negative in 13 analyses. Thus, the decision latencies decrease as a function of the number of arcs between the queried event and the answer event. Causal strength was significant in only 1 out of 15 analyses. Topic familiarity and argument overlap did not have a consistent significant impact on decision latencies. In follow-up multiple regression analyses with interaction terms (AxC, AxD, CxD, and AxCxD), only 4 out of the 60 interaction terms were statistically significant.

The fact that GOA latencies decreased as a function of structural distance is incompatible with a spreading activation explanation of distance effects. A spreading activation explanation would predict a positive relationship between distance and latencies. The structural distance evaluation reflects discrimination processes rather than search processes (Wagener & Wender, 1990). That is, it is more difficult to discriminate the relative temporal order of two events when the two events are structurally close.

Summary comments on expository text studies. These studies showed consistent support for the arc search procedures and structural distance, but very little support for constraint satisfaction (causal strength and argument overlap). The fact that the constraint satisfaction came up empty suggests that the comprehenders must have a deep level of domain knowledge before they can assess causal relationships between events; the college students in this study probably did not have an impressive amount of background knowledge for the scientific mechanisms depicted in the text.

The arc search procedures and structural distance would go a long way in converging on a small number of good answers to a question. Given that the texts in this study had 5 events, there were 4 explicit candidate nodes that were potential answers to each question. On the average, half of these answers would be pruned out by the arc search procedure, which would pursue either causal antecedents or causal consequences but not both. Structural distance would further decrease the space because answer quality decreases exponentially as a function of the number of arcs between the candidate node and the queried node. According to our best-fit estimates of the dampenning curve, 1.3 out of the 4 nodes would be good answers, a convergence ratio of .33. As the database grows in volume and the graph structures have more diverse paths, the convergence ratio gets closer to 0 (Graesser & Franklin, in press).

With one exception, QUEST's arc search procedures are adequately captured in Figure 6. The exception lies in the why questions, which interact with type of knowledge domain. Causal antecedents are prevalent when why-questions are answered in the context of physical systems whereas causal consequences are more prevalent when technological and biological systems are probed. This difference was explained by postulating that a teleological goal hierarchy was superimposed on the networks in biological and technological systems, but not in physical systems.

Question Answering in the Context of Narrative Text

In a series of studies we collected question answering protocols and GOA judgments after college students comprehended simple narrative texts. Unlike the expository texts in the above studies, college students generate a large volume of knowledge-based inferences when they comprehend simple stories and scripts. According to some estimates, the volume of knowledge-based inferences is 4-5 times greater in narrative text than expository text (Graesser, 1981; Graesser & Clark, 1985). The actions and events depicted in narrative have a close correspondence to mundane everyday experiences so inference mechanisms are more automatic. Narrative is an excellent genre to study because there is an extensive interplay between episodic knowledge and generic knowledge.

There has been a great deal of support for QUEST's convergence mechanisms in studies that have collected question answering protocols after subjects comprehend narratives (Graesser, 1978, 1981; Graesser & Clark, 1985; Graesser & Murachver, 1985; Graesser et al., 1980; Graesser et al., 1981). Answer production scores for why, how, when, where, enable, and CONS questions are significantly predicted by QUESTs arc search procedures, by structural distance, and by many dimensions of constraint satisfaction. These findings were reported in studies published before this contract, so they will not be covered in this report. The research conducted in this contract focused on GOA judgments for question-answer pairs after subjects comprehended short narrative passages. This research has been written up in two studies (Golding, Graesser, & Millis, in press; Graesser, Lang, & Roberts, 1989).

College students at Memphis State University first read one of two stories and then judged the quality of particular answers to particular questions about actions and events in the story. One of the stories is provided below, followed by two example question-answer pairs.

The Czar and his Daughters

Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off, they cried for help. Three heroes heard the cries and set off to rescue the daughters. The heroes came, fought the dragon and rescued the maidens. Then the heroes returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heroes.

Why did the heroes fight the dragon?
The dragon kidnapped the daughters.

Why did the heroes fight the dragon?
The daughters were frightened.

The first answer to the question refers to an action that is explicitly stated in the text whereas the second answer is a knowledge-based inference. Our goal was to test QUEST's ability to explain GOA judgments and decision latencies both for answers that were inferences and for answers that were explicit text statements.

The procedure for collecting GOA judgments and latencies was exactly the same as that described for the study on expository text. On each trial the subjects read the question at their own pace by pressing a button. After a brief .5 second delay, the screen was erased and replaced with the answer. The subject indicated their GOA judgment by pressing one of two buttons (BAD answer versus GOOD answer). In some studies we collected discrete GOOD/BAD judgments whereas in others we collected GOA judgments on the following 4-point scale: (1) bad answer, (2) possibly an acceptible answer, (3) moderately good answer, and (4) very good answer. Latencies were not analyzed when judgments were collected on the 4-point scale. Fifty subjects provided binary GOA judgments whereas 60 provided GOA ratings. An equal number of subjects were assigned to each of the five question conditions.

Materials. The narrative texts were two short stories that have been investigated extensively in previous studies (The Czar story and a story about an ant and a dove). Five intentional actions and 4 events were selected as queried nodes from each passage, yielding 18 queried nodes altogether. For example, the queried actions selected from the Czar story were: The daughters walked in the woods, the dragon kidnapped the daughters, the dragon dragged off the daughters, the heroes fought the dragon, and the heroes returned the daughters to the palace.

Sixteen answer nodes were associated with each of the 18 queried nodes, yielding a total of 288 question-answer items. The same 288 items were collected for each of five question categories: why, how, when, enable, and CONS. The wording of the questions and categories varied somewhat among the question categories. For example, if the queried node was The heroes fought the dragon, then the five questions would be: Why/how/when did the heroes fight the dragon?, What enabled the heroes to fight the dragon?, and What are the consequences of the hero fighting the dragon?.

Most of the answers to the questions were inferences (82%) as opposed to explicit statements in the passages. The inferences were sampled from question answering protocols collected by Graesser and Murachver (1985). In that study, Q/A protocols were collected for each explicit statement in the two stories. Each passage statement was probed with the five question categories (why, how, when, enable, CONS); 10 subjects generated answers for each of the five question categories. Associated with each particular question (e.g., Why did the heroes fight the dragon?) was an answer distribution which included all statement nodes that were produced as answers by 2 or more out of the 10 subjects. Associated with each answer was an answer production score.

Our method of sampling nodes from the answer distribution was stratified and random, with a number of criteria that needed to be met before a node was accepted in the answer sample. Regarding stratification, answers were selected from the answer distributions of all five question categories. Regarding randomization, we sampled nodes randomly from the answer distributions of statement N-2, N-1, N, N+1, and N+2, such that N refers to the queried node. The answer distributions associated with passage statement N were weighted higher than the other positions. A node was eliminated from the sample if it was a style specification (e.g., X occurred quickly) or a time index (e.g., in the afternoon, yesterday). We focused on answers that were events, states, actions, and goals because nodes in these categories can be articulated as complete sentences.

When the answers were prepared, the events and states were articulated in exactly the same way across the five question categories. Events and states were declarative sentences in the past tense, with no connectives preceding the statement (e.g., The daughters were frightened, It was a nice day). However, there were fluctuations among question categories as to whether an answer was articulated as an action (e.g., The dragon kidnapped the daughters) or a goal (The dragon wanted to kidnap the daughters). The rules for articulating the answers uniformly gave an answer its best shot at being judged as a good answer to a question (Graesser, Lang, & Roberts, 1989).

Variables and multiple regression analyses. There were three dependent measures: GOA rating (on the 4-point scale), GOA judgment (the binary GOOD/BAD decision), and GOA judgment latency. Three separate sets of multiple regression analyses were performed, corresponding to these three dependent measures. Whenever a multiple regression analysis was performed, variability among question-answer items (averaging over subjects) served as an error term; the beta-weights reported in the subsequent tables are based on these item analyses. However, in all tests of statistical significance, we assessed variability among subjects in addition to variability among items. When variability among subjects was assessed, we performed multiple regression analyses on individual subjects and tested whether the beta-weights of each predictor significantly differed from 0.

The predictor variables in each regression analysis are listed and specified below. Some predictor variables were theoretically interesting from the perspective of the QUEST model, namely those associated with the convergence mechanisms and the information sources.

(1) Arc search. The question-answer item received a score of 1 if there was a legal path of arcs between the entry node and the answer node in the database. The score was 0 if there was no legal path. It should be noted that the textbase contained only explicit passage nodes. When the answer was an inference, it was placed at its "virtual location" in the textbase, as if the inference were an explicit statement. The arc search procedures differed among the five question categories (as discussed in Graesser & Clark, 1985; Graesser & Murachver, 1985), so the values on this variable depended on the question category under consideration. The mean arc search scores were .55, .46, .68, .68, and .40 for the why, how, when, enable, and CONS questions, respectively.

(2) Structural distance. This was the number of arcs between the entry node and the answer node in the textbase. Whenever two nodes were on multiple paths, structural distance was based on the shortest path. The mean score was 1.7 arcs.

(3) Constraint satisfaction. This variable measured the extent to which the answer satisfied a set of semantic and conceptual constraints of the queried node. The values varied from 0 to 1 on each dimension, specifying whether the constraint was not satisfied versus was satisfied, respectively. There were five dimensions: argument overlap, causal strength, temporal compatibility, planning compatibility, and plausibility. These dimensions were defined earlier in the section that described the QUEST model. Two independent raters provided judgments on each dimension and achieved a satisfactory degree of reliability (i.e., between .75 and .96). An overall constraint satisfaction score was computed, consisting of the average of these five dimensions; the mean of this score was .65.

(4) Number of generic information sources. This variable was the number of generic information sources that would supply the answer to the question. Each content word in the queried node served as a potential information source. Decisions needed to be made as to whether a particular answer was stored in a given information source (e.g., whether the node X is frightened is stored in the GKS for FIGHTING). These decisions were based on samples of data collected by Graesser and Clark (1985). Graesser and Clark extracted the content of each GKS associated with the two stories. The content was extracted empirically by a "free generation plus question answering method;" subjects in one group generated lists of typical properties, actions, events, and other nodes in a particular GKS whereas subjects in another group answered questions about the content extracted from the free generation sample. Associated with each GKS was a list of statement nodes that were generated by 2 or more subjects. Regarding the present study, an answer was scored as coming from an information source if it was a member of the Graesser and Clark node list for that information source. An average answer was a member of approximately 1 generic information source (mean equal .92).

(5) Verbatim statement. This variable specified whether the answer was explicitly mentioned in the passage (value = 1) or whether it was an inference (value = 0). The mean was .18.

(6) Answer production score. This was the likelihood that the particular answer would be produced when a particular question is asked in a question answering task. The answer production scores were extracted from the empirical answer distributions collected by Graesser and Murachver (1985). The mean answer production score was .06.

(7) Queried action/event. The queried node was either an intentional action (value = 1) or an event (value = 0).

(8) Story. The Czar story received a value of 1 whereas the Dove story received a value of 0.

Correlation materices were prepared in order to assess whether there was any serious problem of collinearity among predictor variables. A correlation matrix was prepared for each of the five

question categories; each matrix included the above predictor variables (with the overall constraint satisfaction score instead of the breakdown of the five dimensions) and three additional variables (number of content words in answer, number of syllables in answer, and the logarithm of word frequency for content words). Given that each matrix had 11 variables, there were 55 correlations in each matrix and 95 unique correlations among the five matrices. Only 1 of the 95 correlations was greater than .40; there was a positive correlation between number of content words and number of syllables, $r = .68$. Another 8 correlations had absolute values of .31 to .40. Therefore, 91% of the correlations were small or modest.

Goodness-of-answer (GOA) judgments. The mean GOA judgments were .40, .43, .39, .51, and .50 for why, how, when, enable, and CONS questions, respectively. The mean GOA ratings were 1.99, 2.07, 2.51, 2.32, and 2.37. Table 4 presents the outcomes of the multiple regression analyses on the discrete GOA judgments and the GOA ratings. Beta-weights are presented for 8 predictor variables, segregated by question category. When averaging over question categories and considering item variance, the multiple regression equations accounted for 50% of the variance of binary GOA judgments and 52% of the variance of GOA ratings. All of the multiple regression equations significantly predicted the GOA decisions. Table 4 indicates whether each predictor variable was significant in the item analyses and in the subject analyses; we declared a predictor as significant if it was statistically significant in both the item analysis and subject analysis.

The regression equations were almost identical for the binary GOA decisions and the GOA ratings. This can be illustrated by computing the proportion of beta-weights that had the same qualitative outcome, i.e., both were significantly positive, both negative, versus both nonsignificant. Of the 40 beta-weight comparisons in Table 4, 35 had the same qualitative outcome (88%). The values of the beta-weights were also quantitatively similar. When the binary GOA judgments were compared to the ratings, the mean beta-weights (averaging over question category) were virtually identical: arc search (.45 versus .43, respectively), structural distance (-.09, -.09), constraint satisfaction (.23, .23), information sources (.04, .01), verbatim answer (.04, .06), answer production score (.26, .29), queried action/event (.01, -.01), and story (-.02, .05).

Support was found for all three components of the convergence mechanism. Arc search and constraint satisfaction had significantly positive beta's in all 10 multiple regression analyses. Structural distance had significantly negative beta's in 7 out of 10 equations. The when and CONS question did not show consistent support for structural distance. In addition, the bivariate correlations were perfectly compatible with the beta-weights in Table 4.

We performed some follow-up multiple regression analyses that assessed interactions among the three components of convergence: arc search (A), structural distance (D), and constraint satisfaction (C). That is, we added four interaction terms to the multiple regression equation (AxC, AxD, CxD, and AxCxD). The three interaction terms significantly increased the amount of explained variance in the item analyses from 51% to 52%. Two natural groups of questions emerged on the basis of the patterns of 3-way interactions; why, how, and enable questions formed one group whereas the when and CONS questions formed the other. Significant 3-way interactions were found for the first group but not the second group. Figure 9 plots the 3-way interactions for the two groups of subjects, segregating the binary judgments and the ratings. The b-weights (i.e., nonstandardized regression coefficients) of the three main effects and four interaction terms were used to generated the values in Figure 9. As shown in Figure 9, structural distance consistently yielded flat lines for when and CONS questions. For why, how, and enable questions, however, structural distance significantly decreased GOA scores for the legal answers that failed to satisfy constraints and for the illegal answers that satisfied constraints; the other two lines were essentially flat.

The above results failed to support a sequential model which assumes that the arc search procedure is executed before constraint satisfaction. If anything, these two components are performed simultaneously and appear to be additive. Both legal and illegal answers show robust effects of constraint satisfaction. The AxC interaction term was nonsignificant in 7 out of 10 analyses, an outcome that would tend to support an additive model. The GOA judgments were consistently higher for illegal answers that satisfy constraints than for illegal answers that fail to satisfy constraints. The data consistently argue against a sequential model in which the arc search procedure is executed prior to the constraint satisfaction component.

We performed a set of multiple regression analyses which segregated the five dimensions of constraint satisfaction. When averaging across the 5 question categories and two GOA scales, the mean beta-weights were -.02, .10, .03, .14, and .19 for argument overlap, temporal compatibility, planning compatibility, plausibility, and causal strength, respectively. Causal strength was significant in all 10 analyses; temporal compatibility and plausibility were each significant in 7 analyses; planning compatibility was significant in 4 analyses; and argument overlap was significant in only 2 analyses (with one positive and one negative beta). When considering these 30 significant effects, 29 were in the direction that would be predicted by QUEST. Therefore, there was evidence for 4 out of the 5 dimensions of constraint satisfaction.

Multiple regression analyses were performed on those answers that had answer production scores of 0. These answers to a particular question were never generated by subjects who supplied Q/A protocols in the Graesser and Murachver (1985) study. The same regression analyses were performed as those in Table 4 except that answer production scores were dropped (because the value was always 0). All 10 equations siginificantly predicted the GOA judgments, accounting for 39% of the item variance overall. The beta-weights were qualitatively and quantitatively the same as those in Table 4. Of the 70 beta-weights in these analyses, 65 had the same qualitative outcome as those in Table 4. The mean beta-weights of the theoretically interesting predictors were quantitatively similar for (a) the entire answer sample and (b) those answers with an answer production score of 0: Arc search (.44 and .44, respectively), constraint satisfaction (.23, .28), structural distance (-.09, -.09), generic information sources (.03, .02), and verbatim answer (.05, .04).

The multiple regression analyses consistently failed to support the prediction that GOA judgments would increase as a function of number of generic information sources. Only 1 out of the 10 analyses in Table 4 showed a significant effect for this predictor. As discussed earlier, there was some foundation for anticipating a curvilinear relationship between number of generic information sources and GOA. Specifically, answers that come from many information sources are uninformative whereas answers from no information sources are sometimes difficult to interpret. In order to test for a possible curvilinear relationship, we performed 10 regression analyses with two predictors: Number of information sources (I) and I**2. The I**2 term was not significant in any of the analyses. When we restricted our analyses to legal answers that passed the arc search procedure, again there were no significant I**2 terms. Finally, we assessed whether I interacted with any of the other predictor variables and once again came up empty.

Decision latencies for GOA judgments. Mean decision latencies for the binary GOA judgments were 2.39, 2.40, 2.79, 2.51, and 2.69 seconds for why, how, when, enable, and CONS questions, respectively. When we performed multiple regression analyses on these latencies, we included number of content words, number of syllables, and word frequency as predictors in addition to the other 8 predictors in Table 4.

Table 5 presents the beta-weights from the multiple regression analyses, segregated by the 5 question categories. Separate analyses are presented on the complete set of items and those question-answer items with answer production scores of 0. Each of the 10 regression analyses was statistically significant, accounting for a mean item variance of 35%. The beta-weights were

quantitatively similar between the complete item set and the set with answer production scores of 0: Arc search (.04 versus .08), structural distance (-.04, -.03), constraint satisfaction (.02, .00), information sources (.04, .07), verbatim answer (-.07, -.07), queried action/event (.00, .00), story (.17, .13), content words (.22, .22), syllables (.28, .28), and word frequency (.00, .00).

According to the beta-weights in Table 5, legal answers to why, how, and enable questions tended to have longer latencies than the illegal answers. In contrast, the arc search beta-weights for when and CONS questions were either nonsignificant or negative. It should be noted that the same patterns of data occurred for the expository texts reported earlier. The beta-weights for structural distance were negative in 8 out of 10 analyses, but were never significant. The fact that they were negative is entirely consistent with the earlier studies of expository text. The beta-weights for constraint satisfaction were significant in only 1 out of 10 analyses.

We performed follow-up multiple regression analyses that assessed interactions among the three components of convergence (AxC, AxD, CxD, AxCxD). The 3-way interaction was significant in only 2 analyses and these two did not have similar patterns of latencies. Therefore, we performed analyses on 2-way interactions. Structural distance did not interact significantly with the other two variables in any of these analyses. We performed a set of analyses on the AxC interaction term along wit the other 11 predictor variables in Table 5. This AxC interaction was either significant or almost significant for why, how, and enable questions, but not for when and CONS questions. It should be noted that the GOA judgments also manifested natural groupings for why, how, and enable questions versus when and CONS questions.

Figure 10 plots the interaction between arc search and constraint satisfaction, segregating why, how, and enable questions from the when and CONS questions. Each interaction was generated on the basis of the b-weights associated with A, C, and AxC predictors in the equation. For when and CONS questions, arc search and constraint satisfaction had essentially no impact on decision latencies. For the other three question categories, however, the illegal answers that failed to satisfy constraints were much faster than than the other three item categories (i.e., legal answers satisfying constraints, legal answers not satisfying constraints, and illegal answers satisfying constraints); the difference in the latency was .39 second. Another way of viewing these data is that the answers which involved a discrepency between arc search and constraint satisfaction (i.e., legal answers failing constraints and illegal answers satisfying constraints) had longer decision latencies than the answers that either failed or succeeded on both components; this difference was .29 second.

Summary of studies on narrative text. All three components of QUEST's convergence mechanism were supported in our analyses of GOA judgments: Arc search, structural distance, and constraint satisfaction. The patterns of GOA judgments and decision latencies provided some clues about the interaction among the three convergence components. The processing of why, how, and enable questions were somewhat different from that of when and CONS questions (as was the case for the studies on expository text). Consider first the why, how, and enable questions. The arc search and constraint satisfaction components are apparently executed in parallel. When output from the two components were in agreement (i.e., both GOOD or both BAD), then the appropriate decision was made. More time was needed to determine that the answer is good than the answer is bad; good answer require that multiple criteria be satisfied whereas bad answers can be detected as soon as there is a failure on a few criteria. When the output from the arc search component is positive, but the output from the constraint satisfaction component is negative, then additional time was needed to evaluate the answer on structural distance. Subjects apparently used structural distance as a criterion for breaking a tie between the arc search and the constraint satisfaction components.

The structural distance evaluator consists of a process of judging distance rather than a process of searching through a structure. If anything, the decision latencies were faster for answers that were at greater distances from the queried node. Any mechanism that emphasizes search processes, such as spreading activation or marker passing, would have predicted longer latencies for structurally distant answers.

Structural distance did not play a significant role when individuals judged answers to when and CONS questions. The arc search and constraint satisfaction components were processed in parallel, with no speed advantage for good answers over bad answers. In should be noted that Graesser and Murachver (1985) also reported that structural distance plays a minimal role for CONS and when questions in Q/A tasks.

There was no evidence for a sequential order in the processing of arc search and constraint satisfaction. These two components combined in an additive fashion. However, there was evidence for sequential processing when structural distance was analyzed. Structural distance was evaluated after the arc search and constraint satisfaction components were completed (in the case of why, how, and enable questions). That is, arc search and constraint satisfaction preceded structural distance evaluation.

The number of generic information sources had a negligible impact on the GOA judgments and latencies. One reason for this null effect may be that the textbase involved simple stories with very familiar content words. The content words triggered GKSs that were well learned and highly automatized. Perhaps expository texts on difficult unfamiliar topics would show greater effects of multiple information sources.

Question Answering in the Context of Generic Knowledge Structures

We collected GOA judgments and decision latencies for question-answer pairs in the context of generic knowledge structures. College students were tested on approximately 500 trials that had the following phases:

> Generic concept: Consider the concept of HOME
> Read question: How does a person clean the house?
> [Subject presses button followed by .5 second pause].
> Judge answer: The person gets a broom.
> [Subject presses BAD or GOOD answer button]

The subject received different generic concepts from trial to trial. GOA judgments were collected for why, how, enable, and CONS questions in the context of 8 different generic knowledge structures: time, tree, home, child, hero, crying, walking, and fighting. These 8 GKSs cover a broad landscape of concepts: abstract concepts, plants, locations, humans, events, and intentional actions. These concepts are broadly distributed among Keil's ontological categories (Keil, 1979).

The 8 concepts were selected from the 51 GKSs that were investigated by Graesser and Clark (1985). As discussed earlier, Graesser and Clark used a free generation plus question answering method to extract the content of each GKS; one group of subjects supplied free generation protocols whereas a second group answered a why and a how question about each statement node in the free generation set. The total set of nodes included all statements generated by 2 or more subjects. A conceptual graph structure was subsequently prepared for each GKS, using QUEST's rules of composition and the representational system specified by Graesser and Clark. An average GKS contained 166 statement nodes.

Four intentional actions and 4 events were selected as queried nodes in each GKS. The selection of answers was slightly different for queried actions and events. For each queried action, there were 16 answer nodes; these nodes were selected using a a stratified random sampling procedure which insured that approximately half of the answers passed the arc search procedure. Given that there were 8 GKSs, 4 queried actions per structure, and 16 answers per queried node, there were 512 unique pairs of queried node and answer node. These same 512 pairs were tested on each question category (why, how, enable, and CONS). The selection of items was the same for queried events, except that only 8 answers were sampled per queried node (yielding 256 unique pairs of queried node and answer node).

Each subject provided responses for 512 items that spanned all four question categories. Eight subjects received any given question-answer pair. Separate groups of subjects were assigned to queried actions versus queried events. The order in which items were presented and tested was randomly determined for each subject separately. Both the GOA judgment and the decision latency were recorded by the computer on each trial.

Variables in multiple regression analyses. The criterion variables were GOA judgment and decision latency. The predictors of primary theoretical interest were arc search, structural distance, constraint satisfaction (the overall measure, as well as the five dimensions of argument overlap, temporal compatibility, planning compatibility, plausibility, and causal strength). We also scaled each answer on number of information sources by having trained judges assess whether a given answer would be stored under each content word in the queried node. For example, consider the question "How does a person clean the house?" and the answer "the person gets a broom;" the judges assessed whether "X get broom" is stored under PERSON, under CLEAN, and under HOUSE (using a 3-point scale). Answer production scores were computed for each question-answer pair by collecting Q/A protocols from a sample of college students at Memphis State University.

In addition to the above theoretically interesting predictors, there were several predictors that were of less interest. These included the 8 dummy coded variables corresponding to the 8 concepts, and dummy coded variables corresponding to different groups of subjects who received particular item sets. We also included the number of content words in the answer, the mean imagery rating per content word, the mean word frequency per content word (the logarithm), and number of syllables in the answer whenever decision latencies were analyzed; these variables are known to substantially influence reading times (Haberlandt & Graesser, 1985).

GOA judgments. Table 6 presents the outcome of the multiple regression analyses on GOA judgments. GOA judgments were significantly predicted in each of the 8 analyses, accounting for an average of 35% of the item variance. The arc search component was significant in all 8 analyses (mean beta = .31) whereas most of the analyses showed significant effects of structural distance (beta = -.12) and constraint satisfaction (beta = .16). Answers had higher GOA judgments if they matched a node in the verb GKS (mean beta = .13) and in a GKS associated with a noun (beta = .07). Therefore, answer quality increased as a function of the number of information sources supplying the answer. Once again, GOA judgments were predicted by answer production scores (beta = .14).

Analyses were performed on the interaction terms associated with the convergence components in exactly the same way that interactions were analyzed in the study on narrative text. The four interaction terms (AxC, AxD, CxD, and AxCxD) increased the amount of explained item variance from 35% to 37%. The patterns of interactions were quite compatible with our analyses of narrative text (see Figure 9). None of the 8 interaction terms were significant in the case of CONS questions. Regarding why, how, and enable questions, 9 of the 24 possible interaction terms were statistically significant (3 significant effects for the AxD, AxC, and AxCxD interactions). The patterns of the interactions closely replicate the patterns in Figure 9, so we will not plot them in this

report. Structural distance did not influence GOA judgments when arc search and constraint satisfaction were in agreement (i.e., both were GOOD or both were BAD). When there was a disagreement, however, then GOA decreased as a function of structural distance; there was a particularly steep slope for legal answers that failed constraint satisfaction.

We performed some follow-up multiple regression equations that segregated the five dimensions of constraint satisfaction. The mean beta-weights for the five dimensions were: argument overlap (.07), temporal compatibility (.07), planning compatibility (.12), plausibility (.13), and causal strength (.00). The percentage of significant beta's was 42%.

GOA decision latencies. The multiple regression analyses on the decision latencies were not very promising. The arc search component was significant in only 2 out of 8 analyses; the mean beta was -.04, opposite in sign to the beta's in the narrative study. As with all studies in this contract, structural distance had a negative beta in all 8 analyses (mean = -.05) and was significant in 3 analyses. The overall impact of constraint satisfaction on latencies was 0. Verb GKS, Noun GKS, and even answer production scores were rarely significant (3 out of 24 analyses). When we analyzed the interaction terms among the convergence components, the effects were rarely significant (4 of 32 analyses and no agreement in the signs of the significant beta's).

Summary of findings. The GOA judgments showed the same patterns of data for the generic knowledge structures and the narrative passages. There was evidence for all three components of convergence: arc search, structural distance, and constraint satisfaction. These three components showed an interesting interaction in the case of why, how, and enable questions (see Figure 9) but not for CONS questions. Structural distance has an impact when arc search and constraint satisfaction have conflicting output but not when the two components are in agreement. Regarding decision latency, the times decrease as a function of structural distance, in support of a comparative judgment mechanism rather than spreading activation. Otherwise, the latency data were not particularly interesting for GKSs.

Question Answering in Naturalistic Contexts

A series of studies tested QUEST's arc search procedures in the context of naturalistic conversation and complex pragmatic environments (Graesser, Roberts, & Hackett-Renner, in press). In the previous experiments discussed in this report, the pragmatic context was restricted and perhaps unnatural. The texts were short, unteresting, and pointless. The true questioner was unknown and the motivation of the questions was unclear. The questioner (e.g., experimenter, booklet, computer) did not genuinely seek knowledge from a knowledgeable information source. In order to assess whether QUEST is a general model, we tested QUEST in three different naturalistic contexts:

(1) A telephone survey on historical or current events (e.g., the Titanic sinking, Hinkley attempting to assassinate Reagan).
(2) A business interaction in which a customer asks a clerk a question (e.g., How does a person get a credit card? in a bank).
(3) An interview between an expert and a host on a popular television program or educational film (e.g., Nightline with Ted Koppel).

If the arc search procedures of QUEST can account for a substantial proportion of the answers in these contexts, then we would be impressed with the scope and external validity of QUEST.

There are ample reasons for being skeptical about the external validity of QUEST. Questions and answers are embedded in conversations. The content and constraints of a conversation can potentially transform the "literal meaning" of a question and thereby radically alter appropriate

replies. Suppose, for example, that a customer visits a used car lot, points to a car, and asks the salesperson: Why is this 1985 Buick so expensive? Some replies are presented below.

(1) The engine is in perfect condition.
(2) This 1983 Chevy is in good condition.
(3) Why don't you look at this 1983 Chevy?
(4) What price range are we looking at here?

Reply 1 would be accepted by QUEST; it specifies a causal antecedent to the Buick being expensive. Replies 2, 3, and 4 would be reasonable replies to the question in the conversation, but these replies are not accomodated by QUEST. Reply 2 does not address the question. The salesperson inferred, by virtue of the customer's question, that the customer couldn't afford the Buick so the salesperson recommended a less expensive car. Reply 3 is syntactically a question but functionally a directive; neither of these speech act categories are accomodated by QUEST. Reply 4 is a question, both syntactically and functionally, and therefore is beyond the scope of QUEST.

Answers to questions in naturalistic conversations are constrained by the speech participants' goals, plans, and common ground. These pragmatic components were discussed earlier when the QUEST model was articulated. To the extent that the goals and plans become more complex, questions are less likely to be simple information seeking utterances. For example, the questions may be directives, indirect requests, conversation monitors, or rhetorical devices. To the extent that the common ground between speech participants is very high, one would expect more violations of QUEST and perhaps more humorous or sarcastic replies.

The above considerations illustrate some potential limits of QUEST but do not imply that the model is useless. QUEST would be quite useful if it could account for 80% of the answers in naturalistic conversations, with the other 20% being explained by components at the level of conversational meaning.

We analyzed the replies that individuals gave to why, how, when, and CONS questions in the three pragmatically complex environments. We classified each statement node in a reply into one of 17 answer categories. Two trained judges reliably categorized the statement nodes (with reliability scores of .85 or higher among the three studies). Nine categories involved "primary topic information" whereas 8 categories involved "pragmatic" information. The primary topic information consisted of the events and activities referenced by the question (i.e., the event of the Titanic sinking, the procedure of obtaining a credit card). In contrast, the pragmatic information addresses (a) the answerers attitude and reactions to the primary topic information and (b) the social, communicative interaction between the questioner and answerer.

If the answer was primary topic information, it was assigned to one of the following 9 categories: causal antecedent, causal consequence, causal antecedent of causal consequence, superordinate goal/action, subordinate goal/action, consequence of subordinate action, time index, location index, and style specification. For any given question category, only a subset of these categories constitutes legal answers. The legal answers to the questions are specified in Table 7. Only 38% of the cells in Table 7 are legal.

We adopted D'Andrade and Wish's (1985) taxonomy of speech acts in our analysis of the pragmatic information. Their classification scheme not only has a solid theoretical foundation in speech act theory, but also can be used reliably by trained judges. D'Andrade and Wish's scheme has seven categories: Assertion, question, request/directive, reaction, expressive evaluation, commitment, and declaration. We included one additional pragmatic category, called "support information," which supports or motivates the speech acts in the above pragmatic categories. It should be noted that a statement node was assigned to one of these pragmatic categories only if

it did not refer to the primary topic information. All of the answers in the pragmatic categories were considered violations of the QUEST model. Therefore, only 18% of the 17 categories would be accomodated by the arc search procedures of QUEST.

Telephone survey study. Experimenters telephoned citizens in the Memphis community, introduced themselves as researchers at Memphis State University, and asked them one question about a historical event or action. The question categories were why, how, when, and CONS questions. Therefore, the following four questions would be generated from the event The Titanic sank: Why did the Titanic sink?, How did the Titanic sink?, When did the Titanic sink?, and What were the consequences of the Titanic sinking?

Twelve events and intentional actions were queried with these four types of questions: The Titanic sank, the space shuttle Challenger blew up, Governor Dukakis lost the 1988 presidential election, the US inflation dropped, Hinkley attempted to assassinate President Reagan, whale-hunters joined in an effort to rescue the Alaskan whales, Gorbachev instituted an open door policy in Russia, Iran's Ayatolah released the US embassy hostages, the US pulled out of Vietnam, the Olympic games were established, Nixon got involved with Watergate, and Bush chose Dan Quayle as his presidential running mate. Given that there were 12 queried events/actions and four question categories, there were 48 unique questions altogether.

The questioners were 7 research assistants at MSU. The questioners introduced themselves and provided some background information before they asked the question. The questioners stated that they were researchers at MSU, that they were conducting a brief survey on current events and historical events, and that they had only one question to ask. The questioners asked whether the answerer would be willing to complete the survey. If the answerer complied, then the questioner asked the question and the answer was taperecorded. Each questioner collected one observation for each of the 48 questions. The answerers were 336 citizens in the Memphis area who answered the telephone and supplied cooperative answers. Answers were deleted and replaced if the person hung up the telephone or answered "I don't know."

The answers were segregated into statement units according to Graesser and Clark's (1985) representational system. Each statement node was then assigned to one of the 17 answer categories specified above. The number of statement units produced by why, how, when, and CONS questions was 119, 141, 90, and 134, respectively. The total number was 484, or 1.4 statement node per reply.

When considering all four question categories, 94% of the answers referred to primary topic information whereas 6% involved pragmatic information. Within the categories of primary topic information, very few of the answers violated QUEST's arc search procedures. The percentages of violations (within primary topic information) were only 5%, 3%, 2%, and 7% for why, how, when, and CONS questions, respectively. None of these percentages significantly differed from 0. The answers in the pragmatic categories are considered QUEST violations in addition to the violations within primary topic information. When all violations are considered, the percentages of violations were 11%, 8%, 6%, and 15% for why, how, when, and CONS questions. In summary, only 10% of the answers violated QUEST's arc search procedures; only 4% of the answers referring to primary topic information were QUEST violations.

The telephone survey context has a number of distinctive pragmatic assumptions that must be considered. The questions are not genuine information seeking questions in the sense that the questioner is seeking information from a topic expert. The questioners are not asking questions in order to fill gaps in their knowledge base about the Titanic. Instead the questioners are seeking information about the beliefs, attitudes, and expectations of the general public. They are testing the answerers about their knowledge of historical events, in a similar fashion that teachers quiz students. That is, the relevant pragmatic mode is "make me know that you know" rather than the

pragmatic mode of a genuine information seeking question (i.e., "make me know what I don't know").

Business transactions. College students pretended they were customers and asked clerks questions at local businesses. For example, a questioner walked into a bank and asked a teller How does a person get a credit card? The teller's answer was tape-recorded and later transcribed. Compared to the telephone survey, the business context provided a more appropriate foundation for genuine information seeking questions. The questioner went to an expert on the topic under the guise of needing information.

The questioners were 24 students in a research methods course at Memphis State University. The answerers were 144 clerks and other personnel at local businesses in the Memphis community. Each questioner collected 6 Q/A protocols. Altogether, there were 9 questions, each of which was asked in a particular setting with appropriate props. The questions are listed below, with settings in parentheses.

Why are people getting compact disk players? (stereo store)
Why would a person buy expensive sneakers? (shoe store)
Why would a person take vitamin E? (pharmacy)
How does a person get this credit card? (bank)
How would a child play with this toy? (toy store)
How do you cook rice? (supermarket)
What would happen if I put this cake in the freezer? (bakery)
What would happen if I wore another person's glasses for a week? (eye doctor's office)
What would happen if I wore this at a wedding? (clothing store)

We manipulated the context that preceded the question. In half the observations, there was no context except for the expression "Excuse me." In the other half of the observations, the question was preceded by one or two context sentences that clarified the questioner's motives for asking the question. However, the data analyses showed no differences between the two context conditions so the data are collapsed in this report.

The Q/A protocols were analyzed in the same way as in the telephone survey study. The mean number of statement nodes per answer was 5.1, 4.3, and 3.7 for the why, how, and CONS questions, respectively. The vast majority of the 634 statement nodes were from primary topic information (78%) rather than pragmatic information (22%). The percentages of answers in pragmatic categories varied among question categories, with means of 23%, 12%, and 34% for why, how, and CONS questions, respectively. Nearly all of the answers within the primary topic information were consistent with QUEST. The percentages of QUEST violations were 3%, 0%, and 8% for why, how, and CONS questions.

Filmed interviews. Some of the filmed interviews were of experts on topics in science (i.e., Conversations with David Myers, the Brain Series, and the Human Animal Series). The other interviews were of popular or controversial individuals on television programs (i.e., the Phil Donohue Show, Nightline with Ted Koppel, and the McNeil/Lehrer News Hour). The fact that these interviews are filmed adds an important level of pragmatic complexity. In particular, when a person answers a question, there are two classes of listeners that must be taken into consideration. First, there is the interviewer, the person who asks the question. Second, there is the audience who views the film. In this sense, two dialogues are actually being held. This added complexity would presumably increase the number of goals and constraints in the goal structure of the person being interviewed. Given the added complexity in these media events, one might expect the QUEST model to be challenged more severely.

We analyzed the questions and answers in 12 hours of the above filmed interviews. A total of 436 questions were recorded during the interviews. There were 28 why questions and 41 how questions, whereas the when and CONS questions were very low in frequency. Therefore, we restricted our test of QUEST to the why and how questions. The overall number of statement nodes in the answers was 94 for why questions and 143 for how questions. The mean number of nodes per question was 3.6 and 3.3 for why and how questions, respectively.

The answers were analyzed in the same way as in the telephone survey and business transaction studies. The vast majority of answers consisted of primary topic information (75%) rather than pragmatic information (25%). Analyses of the primary topic information revealed that the percentages of QUEST violations were 17% and 4% for why and how questions, respectively. When considering all answers, 68% were consistent with QUEST.

Summary of three studies. The three studies robustly supported QUEST's arc search procedures when we consider the primary topic information. Nearly all of these answers (95% across the three studies) were legal answers. Most of QUEST's violations consisted of pragmatic categories outside of the scope of the model. The most frequent pragmatic categories in our Q/A protocols were assertions (which were not part of the primary topic information), counter-questions, requests, directives, and expressive evaluations. The proportion of answers that occurred in the pragmatic categories increased as a function of the pragmatic complexity of the conversational interaction. According to our analysis of the goals and pragmatic constraints, there was the following order of the three contexts on pragmatic complexity: survey < business transaction < interview. The corresponding percentages of answers in the pragmatic categories were 6%, 22%, and 25%.

There are a number of mechanisms that could explain the QUEST violations. One mechanism focuses on the planning and agenda of speech participants. A reply might not answer the immediate question, but instead address the implicit goals of the questioner and answerer. A second mechanism consists of transforming the literal question into a different question that seems more appropriate in the context. A third mechanism involves the common ground between questioner and answerer. When the common ground is extremely high and new information is difficult to come by, the replies might be humorous, sarcastic, or "off the wall." When the common ground approaches zero, one might have trouble formulating any answer (e.g., a stranger approaching you on the street and asking "Why are people getting compact disk players?"). A fourth mechanism addresses whether the question is a genuine information seeking question (see van der Meij, 1987). There should be more QUEST violations to the extent that a query deviates from being a genuine information seeking question.

REFERENCES

Allen, J. (1983). Recognizing intentions from natural language utterances. In M. Brady, & R.C. Berwick (Eds.), Computational models of discourse. Cambridge, MA: MIT Press.

Alterman, R. (1988). Adaptive planning. Cognitive Science, 12, 393-421.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Appelt, D. E. (1985). Planning English referring expressions. Artificial Intelligence, 26(1), 1-33.

Black, J.B., & Bower, G.H. (1980). Story understanding as problem-solving. Poetics, 9, 223-250.

Bower, G.H., Black, J.B., & Turner, T.J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Bruce, B. C. (1982). Natural communication between person and computer. In W. G. Lehnert & M. H. Ringle (Eds.), Strategies for natural language processing. Hillsdale, NJ: Erlbaum.

Clark, H.H., & Clark, E.V. (1977). Psychology and language. New York: Harcourt Brace Jovanovich.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), Elements of discourse understanding. Cambridge, MA: Cambridge University Press.

Collins, A., Warnock, E.H., Aiello, N., & Miller, M.L. (1975). Reasoning from incomplete knowledge. In D.G. Bobrow and A. Collins (Eds.), Representation and understanding. New York: Academic Press.

Dahlgren, K. (1988). Naive semantics for natural language understanding. Boston: Kluwer Academic Publishers.

Dahlgren, K., & McDowell, J. (1989). Knowledge representation for common-sense reasoning with text. Computational Linguistics, 15, 149-170.

D'Andrade, R. G., & Wish, M. (1985). Speech act theory in qualitative research on interpersonal behavior. Discourse Processes, 8, 229-259.

Dyer, M. G. (1983). In-depth understanding: A computer model of integrated processing for narrative comprehension. Cambridge, MA: MIT Press.

Francik, E.P., & Clark, H.H. (1985). How to make requests that overcome obstacles to compliance. Journal of Memory and Language, 24, 560-568.

Glucksburg, S. & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. Journal of Experimental Psychology: Human Learning and Memory, 7, 311-326.

Golding, J.M., Graesser, A.C., & Millis, K.K. (in press). What makes a good answer to a question?: Testing a psychological model of question answering. Discourse Processes.

Goldman, S.R., & Varnhagen, C.K. (1986). Memory for embedded and sequential story structures. Journal of Memory and Language, 25, 401-418.

Graesser, A.C. (1978). How to catch a fish: The memory and representation of common procedures. Discourse Processes, 1, 72-89.

Graesser, A.C. (1981). Prose comprehension beyond the word. New York: Springer-Verlag.

Graesser, A. C., & Black, J. B. (Eds.). (1985). The psychology of questions. Hillsdale, NJ: Erlbaum.

Graesser, A.C., & Clark, L.F. (1985). Structures and procedures of implicit knowledge. Norwood, NJ: Ablex.

Graesser, A.C., & Franklin, S.P. (in press). QUEST: A cognitive model of question answering. Discourse Processes.

Graesser, A. C. & Golding, J. M. (1988). Annotated bibliography on questioning: Contributions from psychology and artificial intelligence. Questioning Exchange, 2, 315-309.

Graesser, A.C., & Hemphill, D.D. (1990). Question answering in the context of scientific mechanisms. Unpublished manuscript, Memphis State University, Psychology Department, Memphis.

Graesser, A.C., Hemphill, D.D., & Brainerd, L.E. (1989). Question answering in the context of causal mechanisms. In Proceedings of the 11th Annual Conference of the Cognitive Science Society. (pp.621-626). Hillsdale, NJ: Erlbaum.

Graesser, A.C., Lang, K.L., & Roberts, R.M. (1989). Question answering in the context of stories. Unpublished manuscript, Memphis State University, Psychology Department, Memphis.

Graesser, A. C, & Murachver, T. (1985). Symbolic procedures of question answering. In A. C. Graesser, & J. B. Black (Eds.), The psychology of questions. Hillsdale, NJ: Erlbaum.

Graesser, A.C., & Nakamura, G.V. (1982). The impact of schemas on comprehension and memory. In G.H.Bower (Ed.), The Psychology of Learning and Motivation, Vol. 16. New York: Academic Press.

Graesser, A. C., Robertson, S. P., & Anderson, P. A. (1981). Incorporating inferences in narrative representations: A study of how and why. Cognitive Psychology, 13, 1-26.

Graesser, A. C., Robertson, S. P., Lovelace, E., & Swinehart, D. (1980). Answers to why-questions expose the organization of story plot and predict recall of actions. Journal of Verbal Learning and Verbal Behavior, 19, 110-119.

Haberlandt, K. & Graesser, A.C. (1985). Component processes in text comprehension and some of their interactions. Journal of Experimental Psychology: General, 114, 357-374.

Kaplan, S. J. (1983). Cooperative response from a portable natural language system. In M. Brady and B. Berwick (Eds.), Computational models of discourse. Cambridge, MA: MIT Press.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95(2), 163-182.

Lehnert, W. G. (1978). The process of question answering. Hillsdale, NJ: Erlbaum.

Lehnert, W. G., Dyer, M. G., Johnson, P. N., Young, C. J., & Harley, S. (1983). BORIS--An experiment in in-depth understanding of narratives. Artificial Intelligence, 20, 15-62.

Mandler, J. M. (1984). Stories, scripts and scenes: Aspects of schema theory. Hillsdale, NJ: Erlbaum.

Mann, W. C. & Thompson, S. A. (1986). Relational propositions in discourse. Discourse Processes, 9, 57-90.

McKeown, K. R. (1985). Discourse strategies for generating natural-language text. Artificial Intelligence, 27, 1-41.

Meyer, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton, & J. B. Black (Eds.), Understanding expository text. Hillsdale, NJ: Erlbaum.

Miller, G.A., Gallanter, E., & Pribram, K.H. (1960). Plans and the structure of behavior. New York: Holt, Rinehart, Winston.

Millis, K.K. (1989). The time course of constructing bridging and expectation knowledge-based inferences during the comprehension of expository text. Doctoral dissertation, Memphis State University.

Miyake, N., & Norman, D.A. (1979). To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 18, 357-364.

Newell, A., & Simon, H.A. (1972). Human problem-solving. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. (1973). Memory, knowledge and the answering of questions. In R. Solso (Ed.), Contemporary issues in cognitive psychology. Washington DC: Winston.

Norman, D. A. & Rumelhart, D. E. (1975). Explorations in cognition. San Francisco: Freeman

Piaget, J. (1952). The origins of intelligence in children. New York: Norton.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. Psychological Review, 89, 250-280.

Reder, L. M. (1987). Strategy selection in question answering. Cognitive Psychology, 19, 90-138.

Robertson, S. P., Black, J. B. & Lehnert, W. G. (1985). Misleading question effects as evidence for integrated question understanding and memory search. In A. C. Graesser & J. B. Black (Eds.), The psychology of questions. Hillsdale, NJ: Erlbaum.

Schank, R. C., & Abelson, R. (1977). Scripts, plans, goals, and understanding. Hillsdale, NJ: Erlbaum.

Schank, R. C., & Riesbeck, C. K. (1981). Inside computer understanding. Hillsdale, NJ: Erlbaum.

Schmidt, C.F., Sridharan, N.S., & Goodson, J.L. (1978). The plan recognition problem. An intersection of psychology and artificial intelligence. Artificial Intelligence, 11, 45-83.

Shanon, B. (1983). Answers to where-questions. Discourse Processes, 6, 319-352.

Singer, M. (1986). Answering yes-no questions about causes: Question acts and question categories. Memory & Cognition, 14, 55-63.

Singer, M. (in press). The psychology of questions: Answering questions about discourse. Discourse Processes.

Sleeman, D.H., & Brown, J.S. (Eds.) (1982). Intelligent tutoring systems. New York: Academic Press.

Souther, A., Acker, L., Lester, J., & Porter, B.(1989). Using view types to generate explanations in intelligent tutoring systems. In Proceedings of the Eleventh Annual Conference of the Cognitive Science Society (pp. 123-130). Hillsdale, New Jersey: Erlbaum.

Sowa, J. F.(1983). Conceptual Structures: information processing in mind and machine. Reading MA: Addison-Wesley.

Stein, N.L., & Glenn, C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), New directions in discourse processing (Vol. 2). Norwood, NJ: Ablex.

Trabasso, T., Stein, N.L., & Johnson, L.R. (1981). Children's knowledge of events: A causal analysis of story structure. In G.H. Bower (Ed.), The Psychology of Learning and Motivation, Volume 15. New York: Academic Press.

Trabasso, T. & van den Broek, P. (1985). Causal thinking and the representation of narrative events. Journal of Memory and Language, 24, 612-630.

Trabasso, T., van den Broek, P. & Lui, L. (1988). A model for generating questions that assess and promote comprehension. Questioning Exchange, 2, 25-38.

van den Broek, P. (1990). Causal inferences and the comprehension of narrative text. In A. C. Graesser, & G. H. Bower (Eds.), The Psychology of Learning and Motivation: Inferences and text comprehension. New York: Academic Press.

van der Meij, H. (1987). Assumptions of information-seeking questions. Questioning Exchange, 1(2), 111-117.

Table 1

Composition Rules for Six Categories of Arcs

A = source node

B = end node

CONSEQUENCE (C)	A causes or enables B	
	A precedes B in time	
	{event | state | style} --C--> {event | state | style}	
IMPLIES (Im)	A implies B	
	A and B overlap in time	
	{event | state | style} --Im--> {event | state | style}	
REASON (R)	B is a reason or motive for A	
	B is a superordinate goal of A	
	(goal) --R--> (goal)	
MANNER (M)	B specifies the manner of accomplishing A	
	A and B overlap in time if the goals are achieved	
	(goal) --M--> (goal | style)	
	(style) --M--> (style)	
OUTCOME (0)	B specifies whether or not the goal in A is accomplished	
	(goal) --O--> {event | state | style}	
INITIATE (I)	A initiates or triggers the goal in B	
	{event | state | style} --I--> (goal)	

Table 2
Beta-weights from Multiple Regression Analyses on Answer Production Scores (Expository Text)

Predictor Variable	Question Category						
	Why			How	Enable	When	CONS
	Physical	Biological	Technological				
Arc search procedure	.44*	.27*	.30*	.37*	.41*	.49*	.57*
Structural distance	-.22*	-.16*	-.12*	-.23*	-.24*	-.39*	-.15*
Causal strength	.07	.00	.02	-.02	.04	.12*	.04
Topic familiarity	-.23*	.02	-.19*	-.04	-.04	-.10*	-.11*
Argument overlap	-.03	-.02	.16*	.05	.01	.04	.09*
Knowledge domain							
Physical				.02	.03	.11*	.16*
Biological				--	--	--	--
Technological				.10*	.07	.09*	.17*
Variance explained in item analysis (R^2)	.30	.10	.15	.21	.23	.42	.40

Note: * significant at $p < .05$ in item analysis

Table 3
Beta-weights from Multiple Regression Analyses on Goodness-of-Answer Judgements and Decision Latencies (Expository Text)

Question Group	GOA Judgements			Decision Latencies		
	Arc Search	Structural Distance	Causal Strength	Arc Search	Structural Distance	Causal Strength
WHY (no connective)						
Physical	.63*	-.35*	.07	.29*	-.13	.05
Biological	.32*	-.39*	.00	.12	-.06	-.17
Technological	.17*	-.52*	.00	.18*	-.21*	-.13
WHY (because connective)						
Physical	.64*	-.46*	.01	.30*	-.18*	-.11
Biological	.42*	-.37*	-.01	.11	-.15*	.01
Technological	.33*	-.53*	-.11	.29*	-.18*	.00
WHY (in order to connective)						
Physical	.45*	-.39*	.16*	.03	.00	-.03
Biological	.69*	-.19*	.00	-.10	-.13	.00
Technological	.69*	-.31*	.16*	.11	.05	.11
HOW	.57*	-.39*	-.04	.33*	-.20*	-.03
ENABLE	.71*	-.29*	.00	.25*	-.10*	-.16*
WHEN (no connective)	.58*	-.43*	.02	.17*	-.08	.00
WHEN (before connective)	.92*	-.02	.00	-.05	-.29*	.01
WHEN (after connective)	.84*	-.28*	.04	.22*	-.12*	.06
CONSequence	.57*	-.42*	.06	.05	-.08	-.04

Note. * significant at $p < .05$ in item analysis

Table 4
Beta Weights of Predictors of Goodness-of-Answer, Segregated by Question Category (Narrative Text)

Predictor Variable	GOA Decision					GOA Rating				
Convergence mechanism	WHY	HOW	WHEN	ENABLE	CONS	WHY	HOW	WHEN	ENABLE	CONS
Arc Search Procedure	$.49_{is}$	$.45_{is}$	$.33_{is}$	$.51_{is}$	$.47_{is}$	$.49_{is}$	$.45_{is}$	$.32_{is}$	$.48_{is}$	$.41_{is}$
Structural Distance	$-.16_{is}$	$-.13_{is}$	$-.06_{s}$	$-.13_{is}$	.03	$-.14_{is}$	$-.15_{is}$	$-.09_{is}$	$-.15_{is}$	.06
Constraint Satisfaction	$.16_{is}$	$.31_{is}$	$.27_{is}$	$.19_{is}$	$.24_{is}$	$.16_{is}$	$.30_{is}$	$.35_{is}$	$.15_{is}$	$.21_{is}$
Information Sources										
Generic Information Sources	$.14_{is}$	.02	$-.03_{s}$	$.06_{s}$	-.01	$.03_{s}$	$.06_{s}$	$-.05_{s}$	$.05_{s}$	-.02
Verbatim/Inference Answer	.01	$.10_{is}$	$.11_{is}$	$.03_{s}$	$.04_{s}$	.01	$.08_{is}$	$.16_{is}$	.02	.02
Auxiliary Variables										
Answer Production Score	$.29_{is}$	$.16_{is}$	$.29_{is}$	$.22_{is}$	$.32_{is}$	$.36_{is}$	$.20_{is}$	$.26_{is}$	$.26_{is}$	$.35_{is}$
Queried Action/Event	$-.07_{is}$	-.02	$.16_{is}$	.02	$-.05_{s}$	$-.10_{is}$	$-.03_{s}$	$.09_{is}$	.02	-.01
Story	$-.08_{is}$	.02	.05	$-.07_{s}$	-.03	$-.06_{s}$	.04	$.09_{is}$	.03	$.16_{is}$
Vararіance explained by multiple regression equation (R^2)	.57	.52	.41	.55	.44	.61	.57	.45	.55	.43

i significant at $p < .05$ in item analysis
s significant at $p < .05$ in subject analysis

Table 5 Beta-weights of Predictors of GOA Decision Latencies (Narrative Text)

Predictor Variables	All Q/A Items					Q/A Items with Zero Answer Production Score				
	Why	How	When	Enable	CONS	Why	How	When	Enable	CONS
Convergence Mechanism										
Arc search procedure	$.16_{is}$	$.09_{is}$	.02	.10	$-.15_{is}$	$.24_{is}$	$.06_{s}$	.09	.12	$-.10_{s}$
Structural distance	$-.05_{s}$	$-.07_{s}$	-.04	-.07	.05	$-.05_{s}$	$-.08_{s}$	-.03	-.06	.05
Constraint satisfaction	$.10_{is}$	.01	.01	-.02	.01	$.08_{s}$	-.04	$.04_{s}$	-.04	-.05
Information Sources										
Generic information sources	.03	$.11_{i}$	.05	.00	.01	$.12_{i}$	.07	$.16_{i}$	.03	-.03
Verbatim/inference answer	-.02	$-.14_{is}$	-.06	$-.09_{i}$	-.04	-.02	$-.14_{is}$	-.06	$-.10_{i}$	$-.04_{s}$
Auxiliary Variables										
Answer production score	$-.27_{is}$	$-.22_{i}$	$-.17_{is}$	$-.16_{is}$	$-.25_{is}$	--	--	--	--	--
Queried action/event	-.07	-.03	.03	$.05_{s}$	.04	-.06	-.03	.04	$.04_{s}$	.03
Story	.07	$.25_{i}$	$.14_{i}$	$.16_{i}$	$.23_{is}$	.05	.18	.19	.10	.14
Content words	$.36_{is}$	$.19_{is}$	$.26_{is}$	$.17_{is}$	$.16_{is}$	$.36_{is}$	$.20_{is}$	$.26_{is}$	$.18_{is}$	$.17_{is}$
Syllables	$.24_{is}$	$.35_{is}$	$.20_{is}$	$.29_{is}$	$.30_{is}$	$.24_{is}$	$.35_{is}$	$.21_{is}$	$.28_{is}$	$.30_{is}$
Word frequency	.02	.00	.05	-.06	.00	.02	.02	.04	-.07	.00
Variance explained by multiple regression equation (R^2)	.42	.38	.29	.26	.38	.43	.38	.30	.27	.38

i significant at $p < .05$ in item analysis
s significant at $p < .05$ in subject analysis

Table 6
Beta weights of Predictors of Goodness-of-Answer Judgements (Generic Concepts)

Predictor Variable	Queried Action				Queried Event			
Convergence Mechanism	WHY	HOW	ENABLE	CONS	WHY	HOW	ENABLE	CONS
Arc Search Procedure	.43*	.42*	.19*	.37*	.36*	.15*	.31*	.27
Structural Distance	-.13*	-.12*	.01	-.14*	-.14*	-.10	-.10	-.26*
Constraint Satisfaction	.21*	.23*	.27*	.25*	.13*	.04	.04	.11
Information Sources								
Verb GKS	.10*	.10*	.15*	.05	.12*	.26*	.22*	.06
Noun GKS	.16*	.05	.03	.07	.04	.13*	.17*	-.07
Auxiliary Variable								
Answer Production Score	.13*	.16*	.20*	.15*	.09*	.18*	.12*	.10
Variance explained by multiple regression equation (R^2)	.42	.46	.31	.34	.35	.30	.36	.25

* significant at $p < .05$ in item analyses

Table 7

Answer Categories Predicted by the QUEST Model.

ANSWER CATEGORY	QUESTION CATEGORY			
Queried Events	WHY	HOW	CONS	WHEN
Causal Antecedent		X	X	
X				
Causal Consequent				X
Causal Antecedent of a Causal Consequent			X	
Time Index		X		X
Location Index		X		
Style Specification		X		
Queried Actions				
Causal Antecedent		X	X	
X				
Causal Consequent				X
Causal Antecedent of a Causal Consequent			X	
Time Index		X		X
Location Index		X		
Style Specification		X		
Superordinate Goal/Action	X		X	X
Subordinate Goal/Action		X		
Consequent of Subordinate Action		X		

The X signifies that the answer category is a legal answer according to the QUEST model.

Figure 1
Components of QUEST: A Model of Human Question Answering

Interpreting the question
Parsing the question into a logical form
Identifying the appropriate question category

Information Sources
Episodic knowledge structures (text experience)
Generic knowledge structures
(concepts, scripts, frames, etc.)
Knowledge is represented as conceptual graph structures

Convergence
Intersection of nodes from different information sources (plus structural distance)
Arc search procedures
Constraint satisfaction

Pragmatics
Goals of questioner and answerer
Common ground
Informativity of answer

Figure 2
An Example Causal Network on Nuclear Power

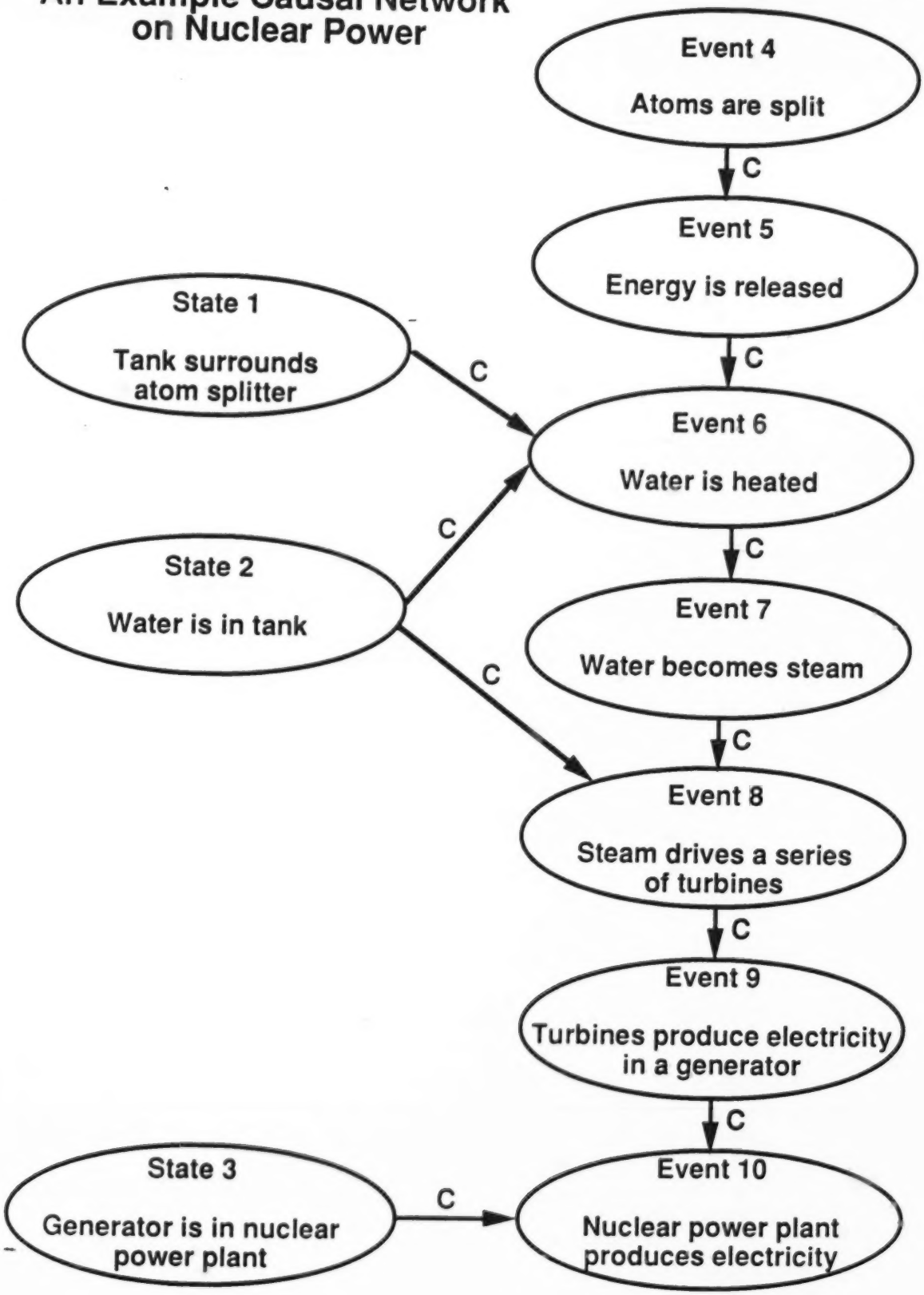


Figure 3
Information Sources for a How-Question

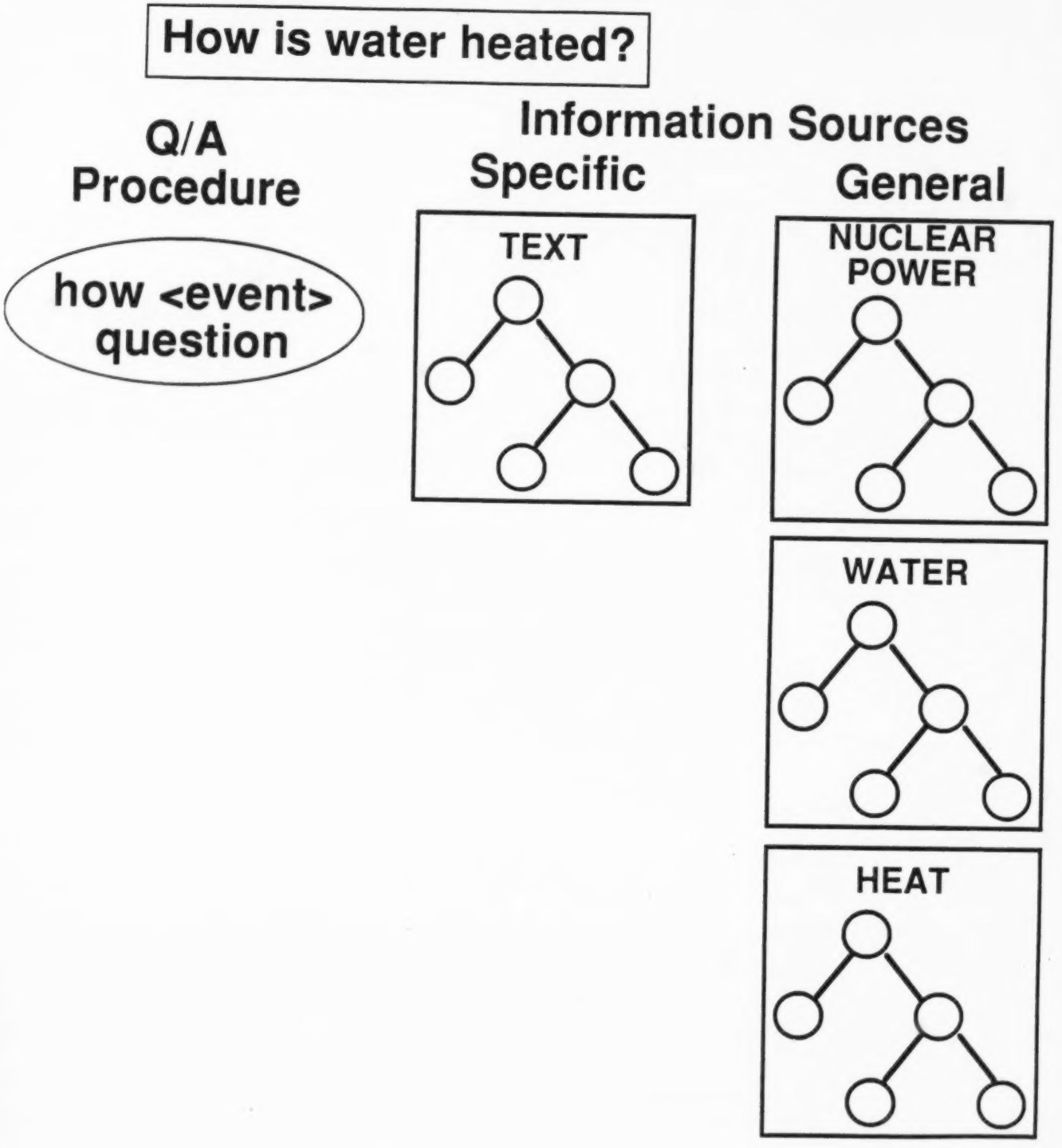


Figure 4
An Example Goal Hierarchy with Goal Initiators

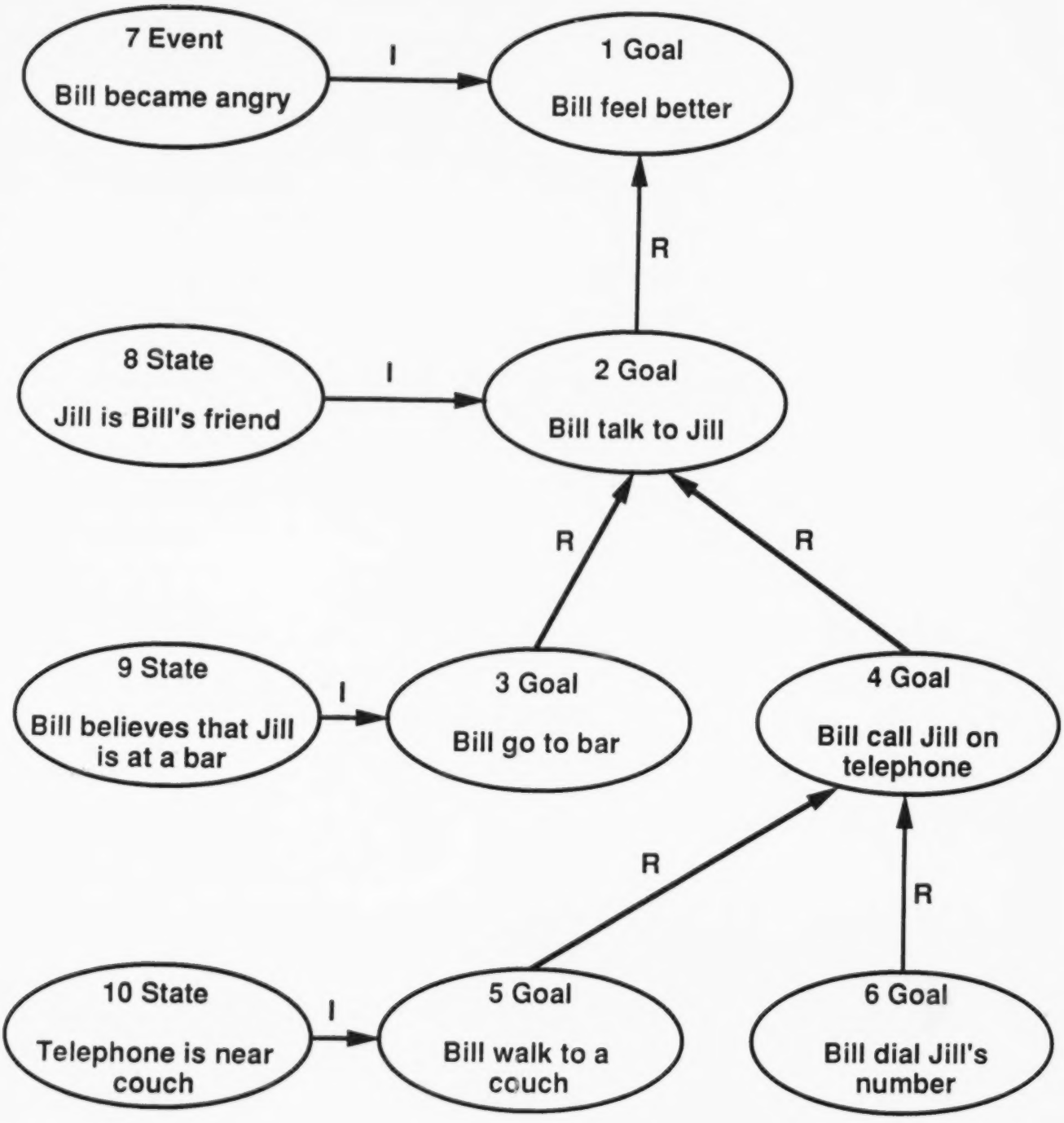
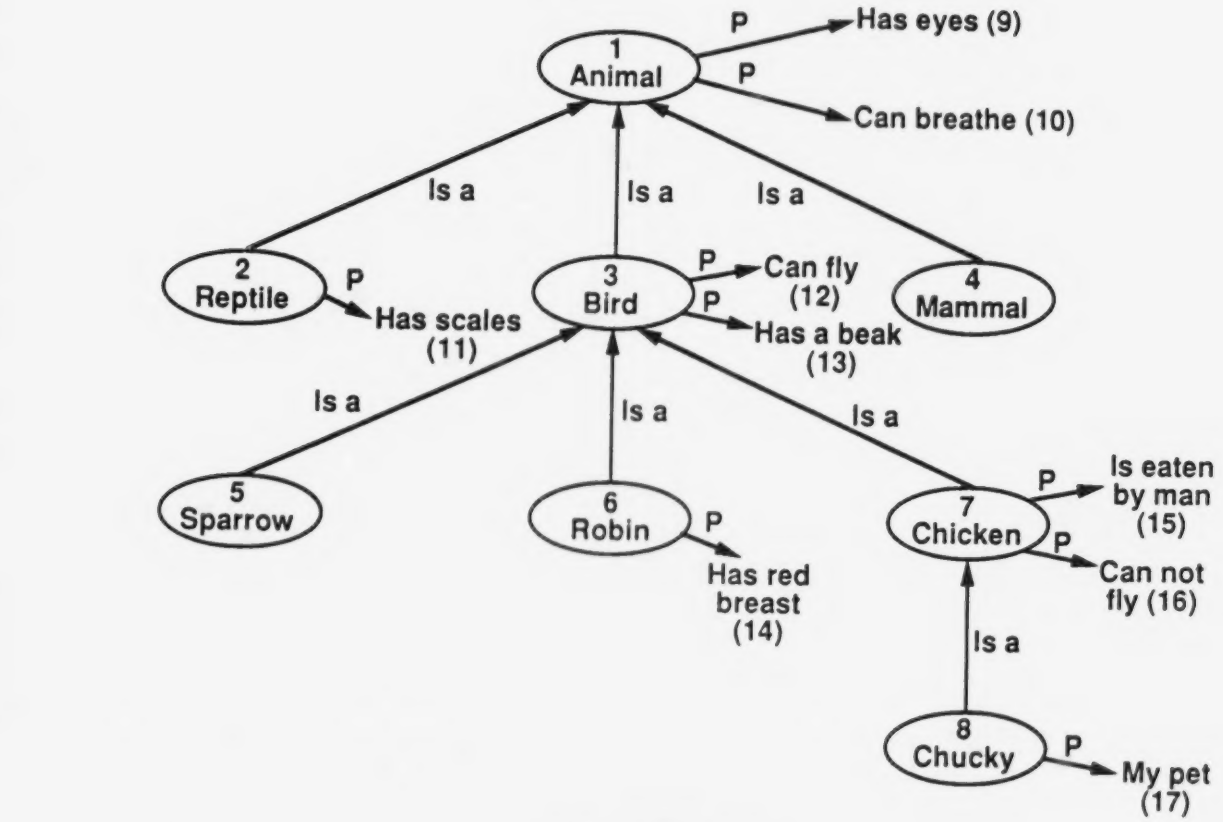


Figure 5
A Taxonomic Structure and a Spatial Partonomy



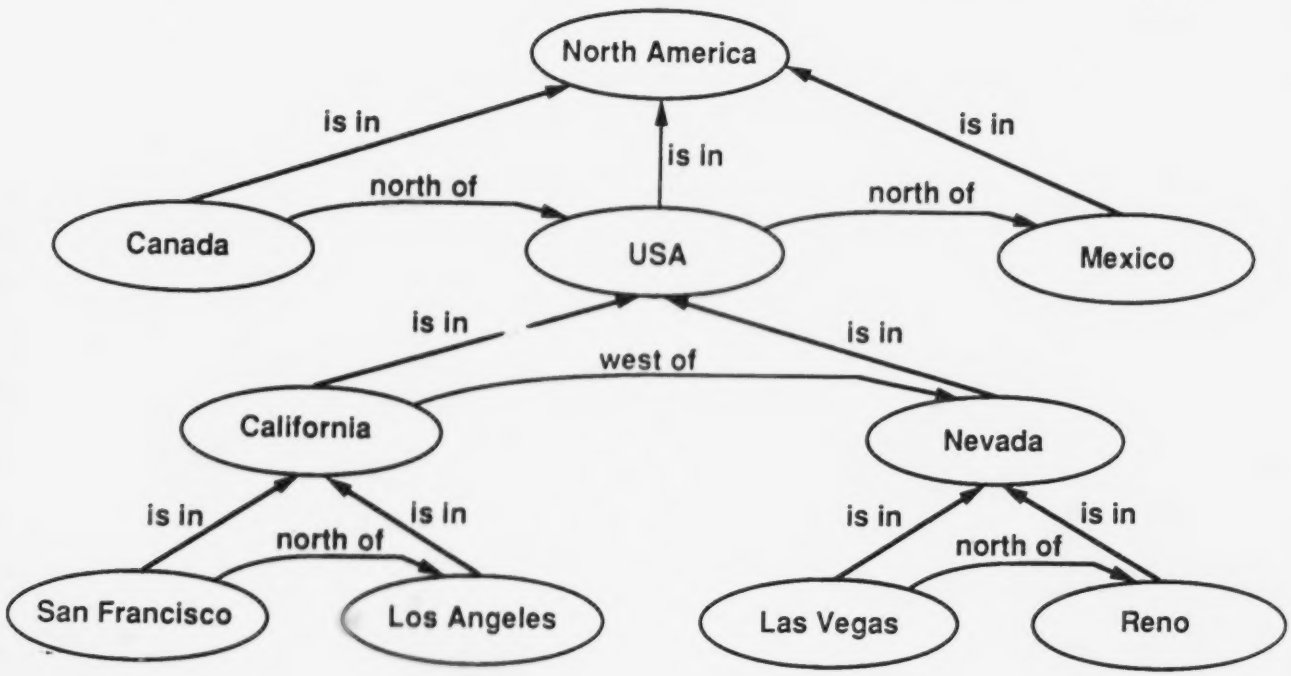
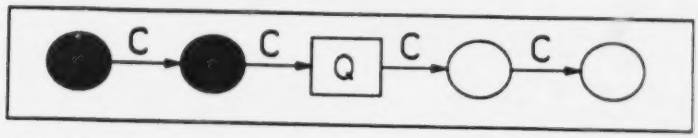


Figure 6
Arc Search Procedures for Events and Actions

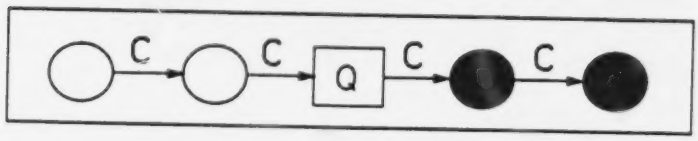
EVENTS

Causal Antecedents



Why?
How?
What enabled?
When?

Causal Consequences

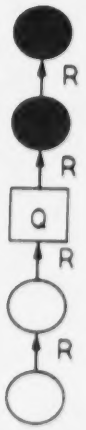


What are the consequences?

ACTIONS

Superordinate Goals

Why?
CONS?



Subordinate Goals/Actions

How?
When?
Enable?

Figure 7
A Goal Structure Running Parallel with a Causal Chain

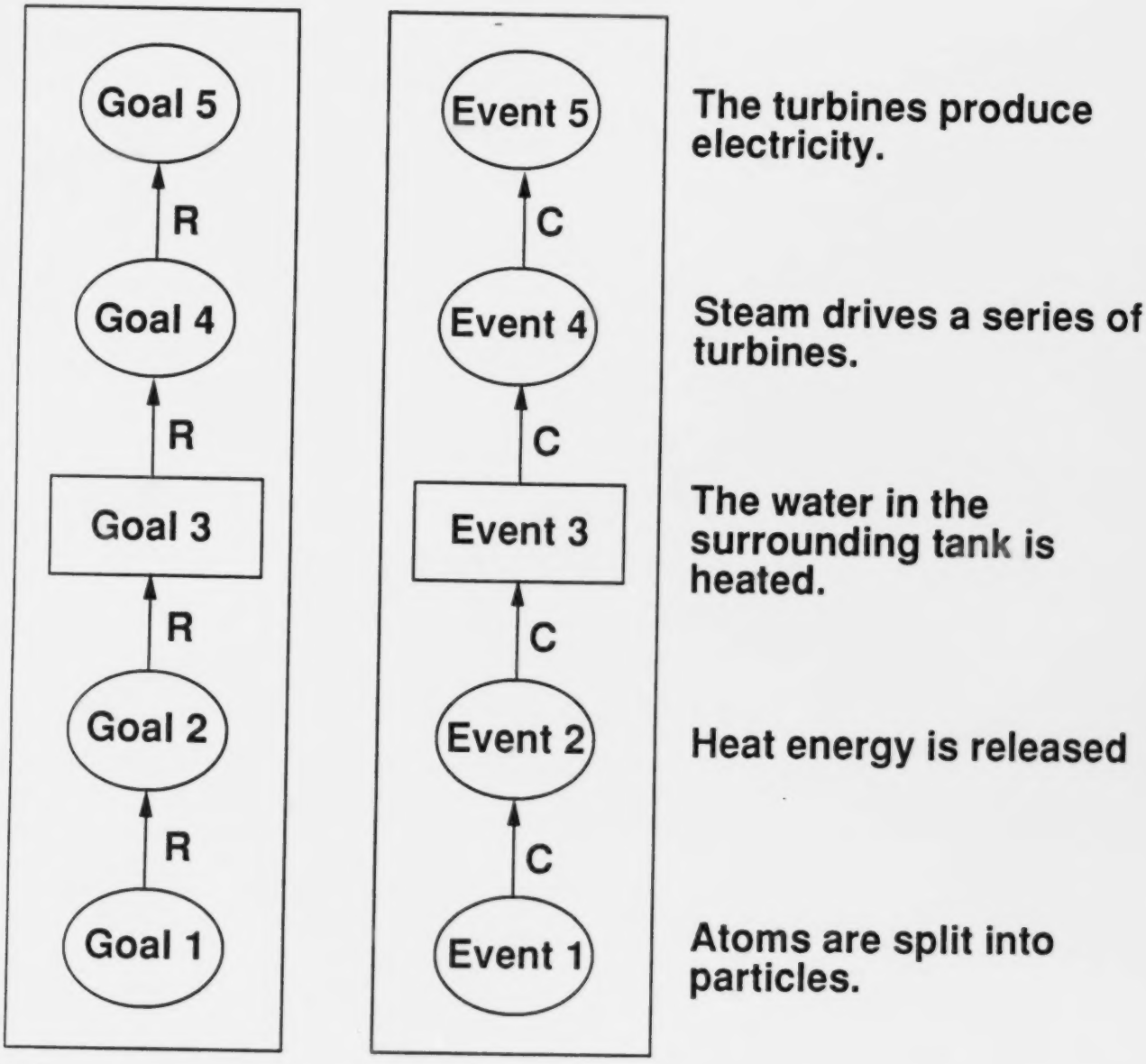


Figure 8
Question-Answer Matrices with Answer Production Scores for Scientific Event Chains

WHY (overall)

Question position \ Answer position	1	2	3	4	5
1	--	.14	.10	.05	.21
2	.17	--	.23	.15	.12
3	.19	.26	--	.21	.15
4	.08	.22	.13	--	.30
5	.08	.13	.11	.23	--

WHY (physical)

Question position \ Answer position	1	2	3	4	5
1	--	.13	.11	.05	.08
2	.34	--	.17	.05	.11
3	.36	.47	--	.08	.09
4	.17	.41	.25	--	.13
5	.14	.25	.19	.38	--

WHY (biological)

Question position \ Answer position	1	2	3	4	5
1	--	.09	.09	.08	.30
2	.08	--	.19	.14	.11
3	.08	.16	--	.31	.22
4	.03	.17	.08	--	.47
5	.04	.06	.11	.19	--

WHY (technological)

Question position \ Answer position	1	2	3	4	5
1	---	.20	.11	.03	.27
2	.08	---	.33	.25	.14
3	.13	.14	---	.23	.14
4	.05	.09	.06	---	.30
5	.05	.06	.05	.14	---

HOW

Question position \ Answer position	1	2	3	4	5
1	--	.13	.05	.06	.06
2	.30	--	.08	.04	.01
3	.22	.31	--	.08	.03
4	.09	.19	.25	--	.04
5	.10	.11	.15	.34	--

ENABLE

Question position \ Answer position	1	2	3	4	5
1	--	.11	.07	.05	.06
2	.36	--	.11	.11	.03
3	.26	.40	--	.12	.04
4	.15	.26	.31	--	.06
5	.10	.21	.18	.34	--

WHEN

Question position \ Answer position	1	2	3	4	5
1	--	.19	.09	.04	.06
2	.62	--	.16	.07	.04
3	.39	.56	--	.19	.05
4	.19	.35	.46	--	.17
5	.13	.16	.25	.45	--

CONS

Question position \ Answer position	1	2	3	4	5
1	--	.35	.30	.21	.30
2	.06	--	.44	.24	.32
3	.04	.13	--	.42	.40
4	.04	.09	.15	--	.56
5	.08	.04	.06	.08	--

Figure 9
Three-way Interaction among Arc Search, Structural Distance, and Constraint Satisfaction (Narrative Text)

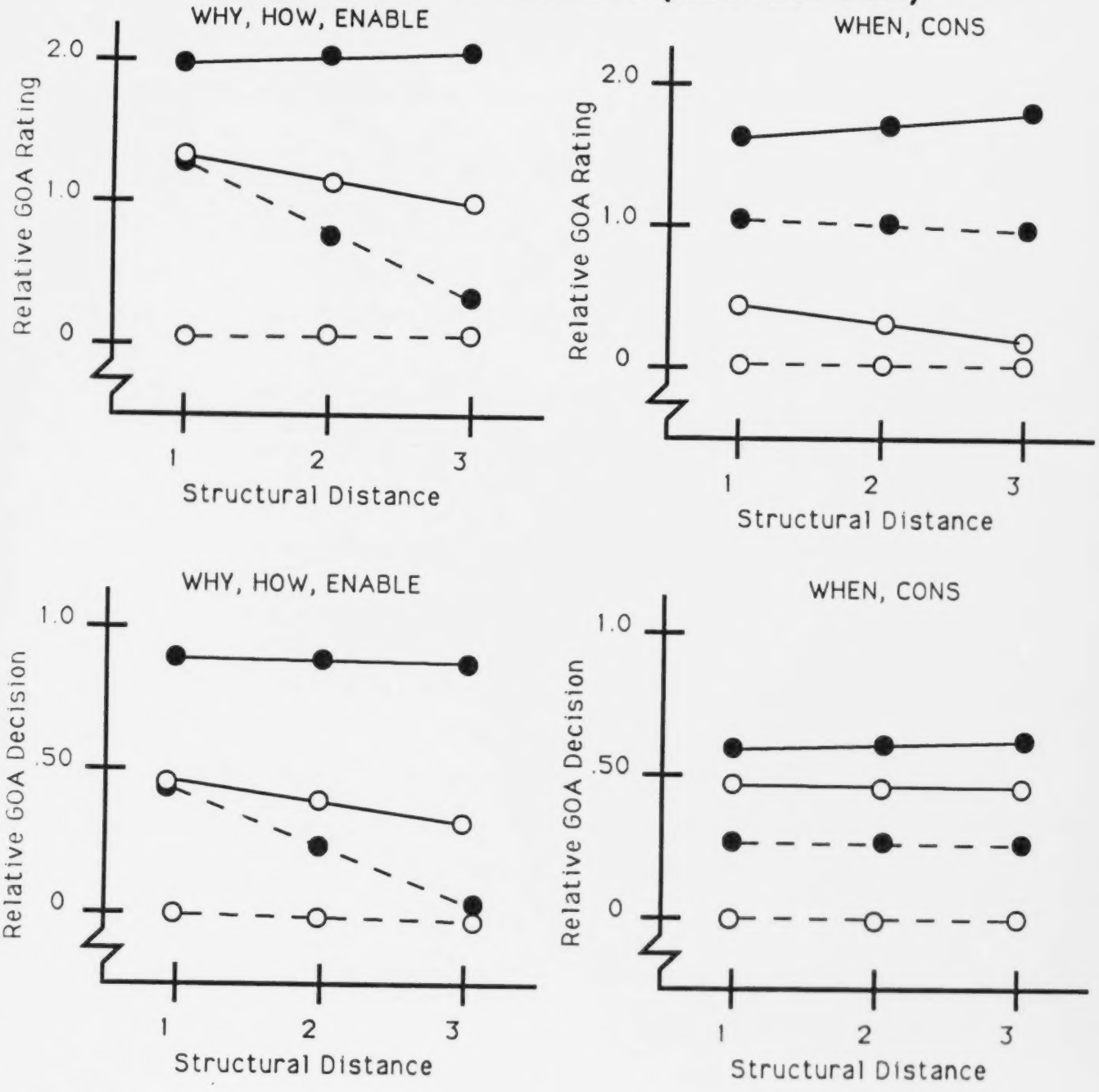
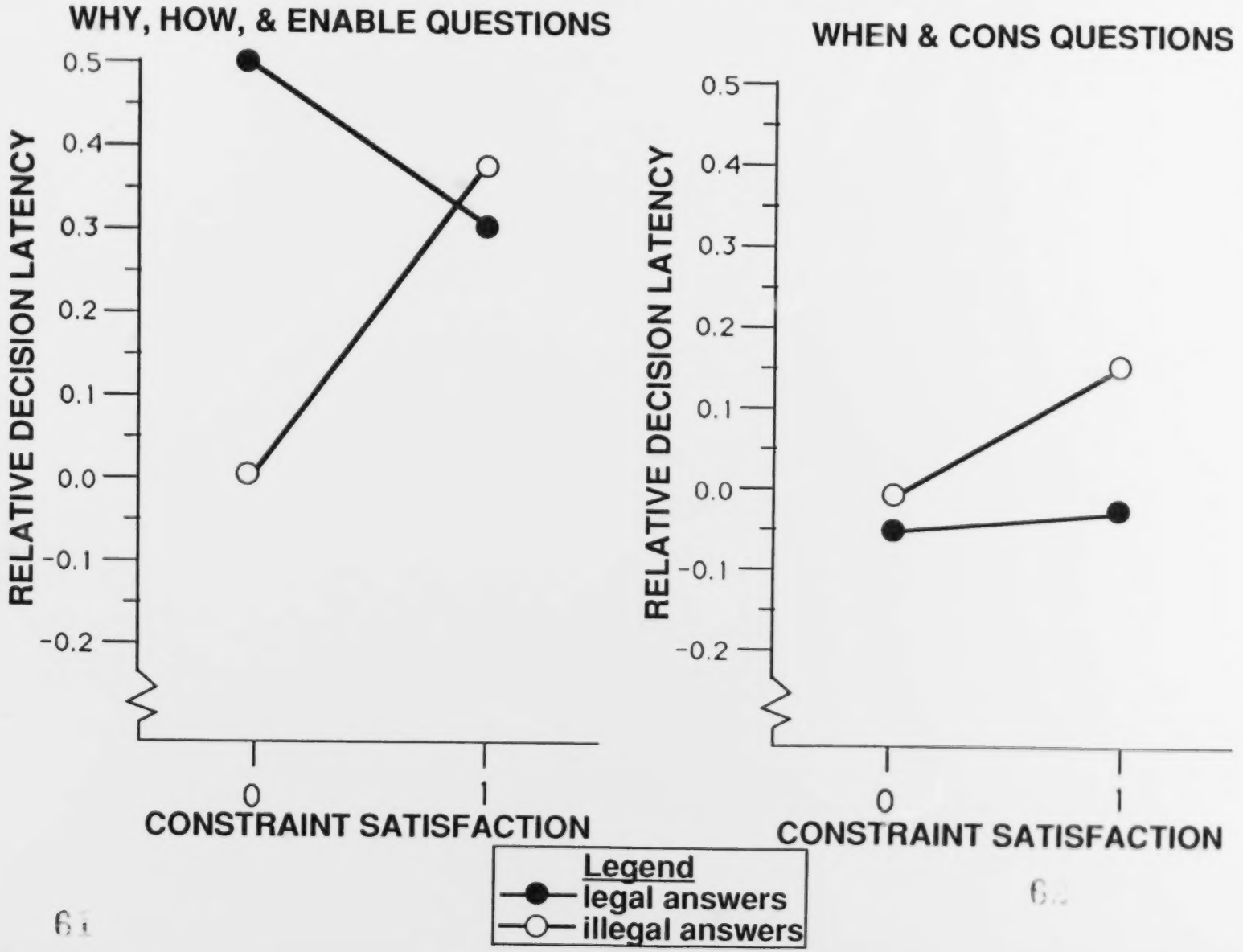


Figure 10
Two-way Interaction Between Arc Search and Constraint Satisfaction for Decision Latencies (Narrative Text)



PUBLICATIONS AND PRESENTATIONS DURING PERIOD OF GRANT

Publications

Dillon, J.T., Golding, J., & Graesser, A.C. (1988). An annotated bibliography of question asking. Questioning Exchange, 2, 81-85.

Graesser, A.C. (1988). Introduction to cognitive science perspectives on question generation. Questioning Exchange, 2, 1-2.

Graesser, A.C. (1988). The first comprehensive text in cognitive science. Contemporary Psychology, 33, 583-585.

Graesser, A.C., & Golding, J. (1988). Models of questioning in cognitive psychology and artificial intelligence. Questioning Exchange, 2, 315-324.

Graesser, A.C., Lang, K., & Horgan, D. (1988). A taxonomy of question generation. Questioning Exchange, 2, 3-16.

Graesser, A.C., Mio, J., & Millis, K. (1988). Metaphors in persuasive communication. In D. Meutsch (Ed.), Models of literary understanding. Amsterdam: Elsevier.

Long, D., & Graesser, A.C. (1988). Wit and humor in discourse processing. Discourse Processes, 11, 35-60.

Long, D.L., Graesser, A.C., & Long, C. (1988). Four computational models for investigating neuropsychological decision-making. In M.A. Williams and C.Long (Eds.), Cognitive approaches to neuropsychology. New York: Plenum.

Graesser, A.C. (1989). The content, sequencing, and timing of skilled action. Contemporary Psychology, 34, 735-738.

Graesser, A.C. (1989). Introduction to creativity in science. In B. Gholson, W. Shadish, R.A. Neimeyer, & A. Houts (Eds.), Psychology of science: Contributions to metascience. Cambridge: Cambridge University Press.

Graesser, A.C., Hemphill, D.D., & Brainerd, L.E. (1989). Question answering in the context of causal mechanisms. In Proceedings of the 11th Annual Conference of the Cognitive Science Society. (pp.621-626). Hillsdale, NJ: Erlbaum.

Graesser, A.C., Long, D., & Mio, J. (1989). Comprehension of humor and wit. Poetics, 17, 86-101.

Graesser, A.C., & Murray, K. (1989). An analysis of verbal protocols to explore the acquisition of a computer environment. In S.P. Robertson, W. Zachary, & J.B. Black (Eds.), Cognition, computing, and interaction. Norwood, NJ: Ablex.

Haberlandt, K., & Graesser, A.C. (1989). Processing of new arguments at clause boundaries. Memory and Cognition, 17, 186-193.

Golding, J.M., Graesser, A.C., & Millis, K.K. (in press). What makes a good answer to a question?: Testing a psychological model of question answering. Discourse Processes.

Graesser, A.C., & Franklin, S.P. (in press). QUEST: A cognitive model of question answering. Discourse Processes.

Graesser, A.C., Golding, J.M., & Long, D.L. (in press). Narrative representation and comprehension. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), Handbook of Reading Research. London: Longman.

Graesser, A.C., & Magliano, J. (in press). Cognition and context. In R. Cohen and A.W. Siegel (Eds.), Context and development. Hillsdale, NJ: Erlbaum.

Graesser, A.C., Magliano, J., & Tidwell, P. (in press). World knowledge, inferences, and questions. (Eds.) Multidisciplinary perspectives on literacy research. An NCTE publication.

Graesser, A.C., Roberts, R.M., & Hackett-Renner, C. (in press). Question answering in the context of telephone surveys, business interactions, and interviews. Discourse Processes.

Haberlandt, K., & Graesser, A.C. (in press). Buffering new information during reading. Discourse Processes.

Haberlandt, K., & Graesser, A.C. (in press). Integration and buffering of new information. In A.C. Graesser and G.H. Bower (Eds.), The Psychology of Learning and Motivation: Inferences and text comprehension. New York: Academic Press.

Haberlandt, K., Graesser, A.C., & Schneider, N.J. (in press). Reading strategies in fast and slow readers. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Lang, K.L., Graesser, A.C., & Hemphill, D.D. (in press). The role of questioning in knowledge engineering and the interface of expert systems. Poetics.

Long, D., Golding, J., Graesser, A.C., & Clark, L.F. (in press). Inference generation during story comprehension: A comparison of goals, events, and states. In A.C. Graesser and G.H. Bower (Eds.), The Psychology of Learning and Motivation: Inferences and text comprehension. New York: Academic Press,

Millis, K.K., Morgan, D., & Graesser, A.C. (in press). The impact of knowledge-based inferences on reading time for expository text. In A.C. Graesser and G.H. Bower (Eds.), The Psychology of Learning and Motivation: Inferences and text comprehension. New York: Academic Press.

Presentations

Long, D., Golding, J., Grassser, A.C., & Clark, L.F. (1988, March). Inference generation during story comprehension: A comparision of goals, events, and states. Paper presented at the Memphis conference on Inference Generation during Text Comprehension, Memphis, TN.

Millis, K., Morgan, D., & Graesser, A.C. (1988, March). The impact of knowledge-based inferences on reading time for expository text. Paper presented at the Memphis conference on Inference Generation during Text Comprehension, Memphis, TN.

Graesser, A.C. (1988, April). Discussant for symposium on "Metaphors in Cognition and Comprehension." The Western Psychological Association, San Francisco, CA.

Graesser, A.C. (1988, April). Models of questioning in cognitive science. Symposium presentation at the American Educational Reasearch Association convention, New Orleans, LA.

Graesser, A.C. (1988, October). Progress report on ONR contract on question answering. Stanford, CA.

Graesser, A.C. & Millis, K. (1988, November). Inference generation during the comprehension of scientific text. Paper presented at the Psychonomics Society, Chicago, IL.

Graesser, A.C. (1989, April). World knowledge, inference generation, and questioning. Invited address at Conference on Reading Research, New Orleans, LA.

Hemphill, D.D. & Graesser, A.C. (1989, June). Testing QUEST with scientific text. Paper presented at American Psychological Society, Arlington, VA.

Graesser, A.C., Lang, K.L., & Roberts, R.M. (1989, July). Question answering in the context of narrative text. Paper presented at the Sylvia Beach Language Comprehension Conference, University of Oregon, Sylvia Beach, Oregon.

Graesser, A.C., Hemphill, D.D., & Brainerd, L.E., (1989, August). Question answering in the context of causal mechanisms. Paper presented at the Cognitive Science Society, Ann Arbor, MI.

Lang, K.L. & Graesser, A.C. (1989, October). Information acquisition strategies of expert computer users when learning new software. Paper presented at the 1989 Human Factors Society Meeting, Denver, Colorado.

Graesser, A.C., & Hemphill, D.D. (1989, November). Answering questions about events in causal mechanisms. Paper presented at The Psychonomic Society, Atlanta, GA.

Graesser, A.C., & Long, D.L. (1989, December). Differences between oral and literary discourse: The effect of pragmatic information on recognition memory. Paper presented at the IGEL Conference, Amsterdam, The Netherlands.

Graesser, A.C., & Millis, K.K. (1990, January). Connectives and coherence in scientific text. Paper presented at the Winter Text Conference, Jackson Hole, WY.

Dr. Terry Ackerman
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Edith Ackermann
Media Laboratory
E15-311
20 Ames Street
Cambridge, MA 02139

Technical Document Center
AFHRL/LRS-TDC
Wright-Patterson AFB
OH 45433-6503

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Mr. Tejwansh S. Anand
Philips Laboratories
345 Scarborough Road
Briarcliff Manor
New York, NY 10520

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Thomas H. Anderson
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champaign, IL 61820

Dr. Stephen J. Andriole, Chairman
Department of Information Systems
and Systems Engineering
George Mason University
4400 University Drive
Fairfax, VA 22030

Prof. John Annett
University of Warwick
Department of Psychology
Coventry CV4 7AL
ENGLAND

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Laura L. Barnes
College of Education
University of Toledo
2801 W. Bancroft Street
Toledo, OH 43606

Dr. William M. Bart
University of Minnesota
Dept. of Educ. Psychology
330 Burton Hall
178 Pillsbury Dr., S.E.
Minneapolis, MN 55455

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Thomas G. Bever
Department of Psychology
University of Rochester
River Station
Rochester, NY 14627

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Gautam Biswas
Department of Computer Science
Box 1683, Station B
Vanderbilt University
Nashville, TN 37235

Dr. John Black
Teachers College, Box 8
Columbia University
525 west 120th Street
New York, NY 10027

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. Deborah A. Boehm-Davis
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Sue Bogner
Army Research Institute
ATTN: PERI-SF
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruynstraat
1120 Brussels, BELGIUM

Dr. Lyle E. Bourne, Jr.
Department of Psychology
Box 345
University of Colorado
Boulder, CO 80309

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. Robert Calfee
School of Education
Stanford University
Stanford, CA 94305

Dr. Joseph C. Campione
Center for the Study of Reading
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Jaime G. Carbonell
Computer Science Department
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. John M. Carroll
IBM Watson Research Center
User Interface Institute
P.O. Box 704
Yorktown Heights, NY 10598

Dr. Robert M. Carroll
Chief of Naval Operations
OP-0182
Washington, DC 20350

Dr. Davida Charney
English Department
Penn State University
University Park, PA 16802

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Mr. Hua Hua Chung
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Dr. Charles Clifton
Tobin Hall
Department of Psychology
University of
Massachusetts
Amherst, MA 01003

Director, Manpower Program
Center for Naval Analyses
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Meredith P. Crawford
3563 Hamlet Place
Chevy Chase, MD 20815

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Ms. Carolyn R. Crone
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Thomas E. DeZern
Project Engineer, AI
General Dynamics
PO Box 748/Mail Zone 2646
Fort Worth, TX 76101

Dr. Ronna Dillon
Department of Guidance and
Educational Psychology
Southern Illinois University
Carbondale, IL 62901

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. J. Stuart Donn
Faculty of Education
University of British Columbia
2125 Main Mall
Vancouver, BC CANADA V6T 1Z5

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
(12 Copies)

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard Duran
Graduate School of Education
University of California
Santa Barbara, CA 93106

Dr. Ralph Dusek
V-P Human Factors
HIL Systems
1225 Jefferson Davis Hwy.
Suite 1209
Arlington, VA 22201

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. John Ellis
Navy Personnel R&D Center
Code 51
San Diego, CA 92252

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

Dr. Susan Epstein
144 S. Mountain Avenue
Montclair, NJ 07042

ERIC Facility-Acquisitions
2440 Research Blvd, Suite 550
Rockville, MD 20850-3238

Dr. K. Anders Ericsson
University of Colorado
Department of Psychology
Campus Box 345
Boulder, CO 80309-0345

Dr. Lorraine D. Eyde
Office of Personnel Management
Office of Examination Development
1900 E St., NW
Washington, DC 20415

Dr. Benjamin A. Fairbank
Operational Technologies Corp.
5825 Callaghan, Suite 225
San Antonio, TX 78228

Dr. Jean-Claude Falmagne
Irvine Research Unit in
Mathematical & Behavioral Sciences
University of California
Irvine, CA 92717

Dr. Marshall J. Farr, Consultant
Cognitive & Instructional Sciences
2520 North Vernon Street
Arlington, VA 22207

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Paul Feltovich
Southern Illinois University
School of Medicine
Medical Education Department
P.O. Box 3926
Springfield, IL 62708

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

CAPT J. Finelli
Commandant (G-PTE)
U.S. Coast Guard
2100 Second St., S.W.
Washington, DC 20593

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Dr. Linda Flower
Carnegie-Mellon University
Department of English
Pittsburgh, PA 15213

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Carl H. Frederiksen
Dept. of Educational Psychology
McGill University
3700 McTavish Street
Montreal, Quebec
CANADA H3A 1Y2

Dr. John R. Frederiksen
BBN Laboratories
10 Moulton Street
Cambridge, MA 02238

Dr. Norman Frederiksen
Educational Testing Service
(05-R)
Princeton, NJ 08541

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Michael Friendly
Psychology Department
York University
Toronto ONT
CANADA M3J 1P3

Col. Dr. Ernst Frise
Heerespsychologischer Dienst
Maria Theresien-Kaserne
1130 Wien
AUSTRIA

Dr. Robert M. Gagne
1456 Mitchell Avenue
Tallahassee, FL 32303

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Drew Gitomer
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O Hara Street
Pittsburgh, PA 15260

Dr. Marvin D. Glock
101 Homestead Terrace
Ithaca, NY 14856

Dr. Sam Glucksberg
Department of Psychology
Princeton University
Princeton, NJ 08540

Dr. Dwight J. Goehring
ARI Field Unit
P.O. Box 5787
Presidio of Monterey, CA 93944-5011

Dr. Joseph Goguen
Computer Science Laboratory
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Dr. Susan R. Goldman
Peabody College, Box 45
Vanderbilt University
Nashville, TN 37203

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Dr. Wayne Gray
Artificial Intelligence Laboratory
NYNEX
500 Westchester Avenue
White Plains, NY 10604

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

H. William Greenup
Dep Asst C/S, Instructional
Management (E03A)
Education Center, MCCDC
Quantico, VA 22134-5050

Michael Haben
DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Bruce W. Hamill
Research Center
The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, MD 20707

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Patrick R. Harrison
Computer Science Department
U.S. Naval Academy
Annapolis, MD 21402-5002

Dr. Frederick Hayes-Roth
Teknowledge
P.O. Box 10119
1850 Embarcadero Rd.
Palo Alto, CA 94303

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. James Hiebert
Department of Educational
Development
University of Delaware
Newark, DE 19716

Dr. Melissa Holland
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Jack Hunter
2122 Coolidge Street
Lansing, MI 48906

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Janet Jackson
Rijksuniversiteit Groningen
Biologisch Centrum, Vleugel D
Kerklaan 30, 9751 NN Haren
The NETHERLANDS

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Daniel B. Jones
U.S. Nuclear Regulatory
Commission
NRR/ILRB
Washington, DC 20555

Dr. Douglas H. Jones
1280 Woodfern Court
Toms River, NJ 08753

Mr. Paul L. Jones
Research Division
Chief of Naval Technical Training
Building East-1
Naval Air Station Memphis
Millington, TN 38054-5056

Dr. Brian Junker
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Ruth Kanfer
University of Minnesota
Department of Psychology
Elliott Hall
75 E. River Road
Minneapolis, MN 55455

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. Frank Keil
Department of Psychology
228 Uris Hall
Cornell University
Ithaca, NY 14850

Dr. Wendy Kellogg
IBM T. J. Watson Research Ctr.
P.O. Box 704
Yorktown Heights, NY 10598

Dr. David Kieras
Technical Communication Program
TIDAL Bldg., 2360 Bonisteel Blvd.
University of Michigan
Ann Arbor, MI 48109-2108

Dr. Thomas Killion
AFHRL/OT
Williams AFB, AZ 85240-6457

Dr. J. Peter Kincaid
Army Research Institute
Orlando Field Unit
c/o PM TRADE-E
Orlando, FL 32813

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80309-0345

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. Richard J. Koubek
Department of Biomedical
& Human Factors
139 Engineering & Math Bldg.
Wright State University
Dayton, OH 45435

Dr. Cary Kress
628 Spazier Avenue
Pacific Grove, CA 93950

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Pat Langley
NASA Ames Research Ctr.
Moffett Field, CA 94035

Dr. Robert W. Lawler
Matthews 118
Purdue University
West Lafayette, IN 47307

Dr. Yuh-Jeng Lee
Department of Computer Science
Code 52
Naval Postgraduate School
Monterey, CA 93943

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. John Levine
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Matt Lewis
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Mr. Rodney Lim
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Charlotte Linde
Structural Semantics
P.O. Box 707
Palo Alto, CA 94320

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. William L. Maloy
Code 04
NETPMSA
Pensacola, FL 32509-5000

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Elaine Marsh
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

Dr. Sandra P. Marshall
Dept. of Psychology
San Diego State University
San Diego, CA 92182

Dr. Clessen J. Martin
Office of Chief of Naval
Operations (OP 13 F)
Navy Annex, Room 2832
Washington, DC 20350

Dr. Manton M. Matthews
Department of Computer Science
University of South Carolina
Columbia, SC 29208

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. James L. McClelland
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Kathleen McKeown
Columbia University
Department of Computer Science
450 Computer Science Building
New York, NY 10027

Dr. Robert McKinley
Law School Admission Services
Box 40
Newtown, PA 18940

Dr. Joseph C. McLachlan
Code 52
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Mr. Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Douglas L. Medin
Department of Psychology
University of Michigan
Ann Arbor, MI 48109

Dr. Jose Mestre
Department of Physics
Hasbrouck Laboratory
University of Massachusetts
Amherst, MA 01003

Dr. D. Michie
The Turing Institute
George House
36 North Hanover Street
Glasgow G1 2AD
UNITED KINGDOM

Dr. George A. Miller
Dept. of Psychology
Green Hall
Princeton University
Princeton, NJ 08540

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Dr. Melvin D. Montemerlo
NASA Headquarters
Code RC
Washington, DC 20546

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. Allen Munro
Behavioral Technology
Laboratories - USC
250 N. Harbor Dr., Suite 309
Redondo Beach, CA 90277

Dr. Ratna Nandakumar
Educational Studies
Willard Hall, Room 213E
University of Delaware
Newark, DE 19716

Dr. T. Niblett
The Turing Institute
George House
36 North Hanover Street
Glasgow G1 2AD
UNITED KINGDOM

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Librarian
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. Paul O'Rorke
Information & Computer Science
University of California, Irvine
Irvine, CA 92717

Dr. Stellan Ohlsson
Learning R & D Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Dr. Judith Reitman Olson
Graduate School of Business
University of Michigan
Ann Arbor, MI 48109-1234

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John Oriel
Navy Training Systems
Center (Code 212)
12350 Research Parkway
Orlando, FL 32826-3224

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Glenn Osga
NOSC, Code 441
San Diego, CA 92152-6800

Dr. Chohoon Park
Army Research Institute
PERI-2
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. Roy Pea
Institute for Research
on Learning
2550 Hanover Street
Palo Alto, CA 94304

Dr. C. Perrino, Chair
Dept. of Psychology
Morgan State University
Cold Spring La.-Hillen Rd.
Baltimore, MD 21239

Dr. Nancy N. Perry
Naval Education and Training
Program Support Activity
Code-047
Building 2435
Pensacola, FL 32509-5000

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Peter Polson
University of Colorado
Department of Psychology
Boulder, CO 80309-0345

Dr. Mary C. Potter
Department of Brain and
Cognitive Sciences
MIT (E-10-039)
Cambridge, MA 02139

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Mr. Steve Reiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Lauren Resnick
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Dr. Edwina L. Rissland
Dept. of Computer and
Information Science
University of Massachusetts
Amherst, MA 01003

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. Ernst Z. Rothkopf
AT&T Bell Laboratories
Room 2D-456
600 Mountain Avenue
Murray Hill, NJ 07974

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
905 Orchid Way
Carlsbad, CA 92009

Nuria Sebastian
Dep. Psicologia Basica
Univ. Barcelona
Adolf Florensa s.n.
08028 Barcelona
SPAIN

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Robin Shealy
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Mr. Colin Sheppard
AXC2 Block 3
Admiralty Research Establishment
Ministry of Defence Portsdown
Portsmouth Hants P064AA
UNITED KINGDOM

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5510
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Zita M. Simutis
Chief, Technologies for Skill
Acquisition and Retention
ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Robert Smillie
Navy Personnel R&D
San Diego, CA 92152-6800

Dr. Edward E. Smith
Department of Psychology
University of Michigan
330 Packard Road
Ann Arbor, MI 48103

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

N. S. Sridharan
FMC Corporation
Box 580
1205 Coleman Avenue
Santa Clara, CA 95052

Dr. Thomas Sticht
Applied Behavioral and
Cognitive Sciences, Inc.
P.O. Box 6640
San Diego, CA 92106

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Patrick Suppes
Stanford University
Institute for Mathematical
Studies in the Social Sciences
Stanford, CA 94305-4115

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-62
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. M. Martin Taylor
DCIEM
Box 2000
Downsview, Ontario
CANADA M3M 3B9

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Thomas J. Thomas
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Perry W. Thorndyke
FMC Corporation
Central Engineering Labs
1205 Coleman Avenue, Box 580
Santa Clara, CA 95052

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Paul T. Twohig
Army Research Institute
ATTN: PERI-RL
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Harold P. Van Cott
Committee on Human Factors
National Academy of Sciences
2101 Constitution Avenue
Washington, DC 20418

Dr. Kurt Van Lehn
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Michael T. Waller
University of Wisconsin-Milwaukee
Educational Psychology Department
Box 413
Milwaukee, WI 53201

Dr. Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
1100 S. Washington
Alexandria, VA 22314

Dr. Diana Wearne
Department of Educational
Development
University of Delaware
Newark, DE 19711

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Federal Aviation Administration
800 Independence Ave, SW
Washington, DC 20591

Dr. Frank B. Withrow
U.S. Department of Education
Room 504D, Capitol Plaza
555 New Jersey Avenue, N.W.
Washington, DC 20208

Mr. Paul T. Wohig
Army Research Institute
5001 Eisenhower Ave.
ATTN: PERI-RL
Alexandria, VA 22333-5600

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Frank R. Yekovich
Dept. of Education
Catholic University
Washington, DC 20064

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611

Dr. Uri Zernik
General Electric:
Research & Development Center
Artificial Intelligence Program
PO Box 8
Schenectady, NY 12301